



MASTERS DISSERTATION

Advanced Analytics and Decision Sciences

Comparison of Sovereign Risk and its Determinants

Anri Smith^{1*}

¹Statistics Department, University of Cape Town, Cape Town, Western Cape, 7701, South Africa

Correspondence

Anri Smith, Statistics Department,
University of Cape Town, Cape Town,
Western Cape, 7701, South Africa
Email: anri7smith7@gmail.com

Funding information

Master program study sponsored by the
National Research Fund (NRF) of South
Africa.

This paper aims to measure, compare and model Sovereign Risk. The risk position of South Africa compared to Emerging Markets as well as in comparison to Developed Markets is considered. Particular interest is taken in how the South African Sovereign Risk environment, and its associated determinants, differs and conforms to that of other Emerging Markets. This effectively highlights how the South African economy is similar to the Emerging Markets and where it behaves differently. Regression, optimisation techniques, dimension reduction techniques as well as Machine Learning techniques, through the use of sentiment analysis, is utilised in this research.

KEYWORDS

Sovereign Risk, Emerging Markets, Developed Markets, Machine Learning, Sentiment Analysis

* Supervisor: Professor Graham Barr

The copyright of this thesis vests in the author. No quotation from it or information derived from it is to be published without full acknowledgement of the source. The thesis is to be used for private study or non-commercial research purposes only.

Published by the University of Cape Town (UCT) in terms of the non-exclusive license granted to UCT by the author.

CONTENTS

1	Introduction	7
1.1	Problem Statement	7
1.2	Problem Context	7
1.2.1	Sovereign Risk	7
1.2.2	Sovereign & Yankee Bonds	8
1.2.3	Sovereign Yield Spread	9
1.2.4	Emerging vs Developed Economies	9
2	Literature Review	11
2.1	Yield Spread	11
2.2	Sentiment Analysis	12
2.3	Regression Yield Spread Modelling	13
2.3.1	Sovereign Yield Spread over the Global Financial Crisis	13
2.3.2	Political Determinants of Sovereign Bond yield spreads	15
3	Data Collection & Exploration	23
3.1	Country Selection	23
3.2	Yield Data	25
3.3	Yield Determinants	26
3.3.1	Economic & Political Data	26
3.3.2	Political News Data	28
3.4	Exploratory Analysis	28
3.4.1	Yield Data	28
3.5	Association Visuals	32
3.5.1	News Data	35

ANRI SMITH	3
4 Methodology	37
4.1 Yield Spread Measure	37
4.2 Sentiment Score Development	38
4.3 Principal Component Analysis (PCA)	40
4.4 Ordinary Least Squares Regression	41
4.4.1 Regression Optimisation Techniques	44
4.4.2 Short-Term Analysis	46
4.4.3 Long-Term Analysis	47
5 Modelling, Results & Analysis	48
5.1 PCA	48
5.2 Regression	50
5.2.1 Short Term Analysis	50
5.2.2 Long Term Analysis	58
5.3 Results Visualisation	60
6 Discussion of Results	61
7 Glossary	62
Appendices	62
A Yield Data Manipulation	62
B Yield Spread Determinants	63
C Predictor Correlations	65
D Error Plots for Short Term Analysis	65
E Stationarity of Errors in Long Term Analysis	66

LIST OF FIGURES

1	Map pinpointing sampled countries across the globe.	24
2	Number of active Sovereign Bonds across different countries as of August 2018.	24
3	Sovereign Bond maturity distribution across different markets for active Sovereign Bonds as of August 2018.	25
4	Yield Transformation Flow Chart.	26
5	South African Position & Comparison to the Emerging & Developed Markets for the period 2009 - 2018 & 2005 - 2018 for 10 year and 30 year bonds respectively.	29
6	South African Sovereign Risk Position within the Emerging Markets for the period 2008 - 2018 & 2005 - 2018 for 10 year and 30 year bonds respectively.	29
7	South African Position within Africa for the period 2008 - 2018 & 2005 - 2018 for 10 year and 30 year bonds respectively.	30
8	South African vs Mexican Sovereign Risk for the period 2008 - 2018 & 2005 - 2018 for 10 year and 30 year bonds respectively.	30
9	South African Sovereign Bond Maturities as indicated by SA Sovereign Yield curves for the period 2014 - 2018.	31
10	Yield Coefficient of Variation for respective South African Sovereign Bond Maturities as indicated by SA Sovereign Yield curves for the period 2014 - 2018.	31
11	Yield summary for the respective markets for the period 2009 - 2018 & 2011 - 2018 for 10 and 30 year bonds respectively.	32
12	Association plots for the respective short term predictor variables.	33
13	Association plots for the respective long term predictor variables.	34
14	Article Collection by Country for the period 2007 - 2018	35
15	Article Collection for the period 2007 - 2018	36
16	Summary of the Sentiment Extraction Process.	38
17	Political Sentiment for the period 2007 - 2018 for South Africa, the United States and the respective markets of interest.	40
18	Short term covariate PCA plot for the period 2012 - 2018	48
19	Long Term covariate PCA plot for the period 2008 - 2016	49

20	Expected Lasso Prediction Error for tested range of lambda values for Emerging Markets yield spread.	54
21	Expected Lasso Prediction Error for tested range of lambda values for Developed Markets yield spread.	55
22	Expected Elastic Net Prediction Error for tested range of lambda values for Emerging Markets yield spread.	56
23	Expected Elastic Net Prediction Error for tested range of lambda values for Developed Markets yield spread.	56
24	Expected Ridge Prediction Error for tested range of lambda values for Emerging Markets yield spread.	57
25	Expected Ridge Prediction Error of tested range of lambda values for Developed Markets yield spread.	57
26	Total reserves (% of total external debt) and the corresponding 30-year yield for the respective Emerging Markets countries.	60
27	Taxes on international trade and the corresponding 30-year yield for the respective Emerging Markets countries.	61
28	Correlation between short term analysis predictors.	65
29	Correlation between long term analysis predictors.	65
30	Short Term Residual Plot for OLS regression on entire set of predictors: Emerging Markets Case	65
31	Long Term Residual Plot for OLS regression on entire set of predictors: Developed Market Case	66

LIST OF TABLES

1	Yield spread determinants according to category	14
2	Political variables that are considered as potential Sovereign Yield spread determinants.	16
3	Findings with regard to political variables	20
4	Findings with regard to country-specific control variables	21
5	Findings with regard to global control variables.	21
6	Findings of interaction regression model	22
7	Summary of Emerging and Developed Market countries used in this research	23
8	Economic Predictors of Interest	27
9	Political Predictors of Interest	28
10	Summary Statistics for the respective sentiment series	40

11	Augmented Dickey Fuller test for short term predictors	43
12	Augmented Dickey Fuller test for long term predictors	43
13	Loading Vectors for the respective Principal Components found for the Short Term Analysis case. . . .	48
14	Loading Vectors for the respective Principal Components found for the Long Term Analysis case. . . .	49
15	Regression Summary Statistics for Emerging Markets yield spread.	51
16	Regression Summary Statistics for Developed Markets yield spread.	52
17	Emerging Markets Case: RSS for each stage of the Forward Stage-wise algorithm.	53
18	Developed Markets Case: RSS for each stage of the Forward Stage-wise algorithm.	53
19	Regression results	58
20	Regression results	59
21	Yield Spread Determinants, their descriptions and data sources	63
22	Regression results	66

1 | INTRODUCTION

1.1 | Problem Statement

This research aims to measure and compare the Sovereign Risk experienced in South Africa in relation to other countries in the respective markets; the Emerging Markets and the Developed Markets. Specifically the positioning of South Africa within the Emerging Markets is considered as well as how its risk profile compares to that of Developed Markets. Furthermore the determinants of Sovereign Risk across the groups of interest are then also explored.

1.2 | Problem Context

In order to effectively compare the Sovereign Risk experienced in South Africa to other countries and their respective markets, Sovereign Risk and how it is measured as well as the characteristics of an Emerging and Developed Market are explored below. Furthermore other factors related to the effective measurement and comparison of Sovereign Risk are also investigated.

1.2.1 | Sovereign Risk

Sovereign Risk refers to the likelihood that a government will default on its debt obligations to banks or other countries. The more likely the probability of default, the more risk faced by the investor who has purchased a government debt security. The term Sovereign Risk is also used to describe government regulation. There may be a lack of regulation or regulation may impede domestic issuers from meeting their debt obligations. Examples of this government regulation include the legal rights of investors as well as foreign currency restrictions [1, 2]. For the purposes of this research Sovereign Risk refers to the debt taken out by government.

The probability that a government will default on its debt is driven by two main factors. These being the government's willingness as well as its ability to service its debt. The government's ability to repay its debt can be measured through economic and socio-political criteria[1].

- The economic component can be expressed as a function of the level of debt and the remaining time-to-maturity of its debt obligations as well as its ability to refinance its debt [1]. Where the refinancing of debt refers to the government taking out an additional loan, in order to settle payment for a maturing loan.
- The socio-political component refers to the country's level of development, social unrest and overall governance[1]. A government's willingness to repay its debt can be estimated based on its historic credit behaviour.

The Sovereign Risk of Emerging Market countries is typically higher when compared to the Sovereign Risk of Developed countries. The probability of default is typically higher in an Emerging economy due to political instability and lack

of government efficiency [3]. However recent protectionist policies introduced by the United States as well as in the United Kingdom, two of the world's most developed economies, may contribute to heightened Sovereign Risk in these countries [2]. The U.S. president, Mr. Trump, has come into office with promises of implementing protectionist policies. This mandate has taken form through the United States-China trade tariffs in July 2018. Where the United States has imposed a wave of tariffs, amounting to billions of dollars, on Chinese goods, which China has responded to in kind [4]. The U.K. on the other hand is in the process of exiting the European Union (BREXIT), giving it the freedom to pursue its own trading policies [5]. The nature of the trading relationship that will ensue between the U.K. and the E.U. is still being determined [5]. Protectionist policies, as mentioned here, have the potential to upset an economy's trade balance and consequentially hinder the government's ability to honour its debt obligations.

The contagion effect of a Global Financial Crisis (GFC) on Emerging economies (BRICS) cannot be confirmed in the early stages of a GFC [6]. This indicates that these economies are initially sheltered from a GFC. However after the Lehman Brothers collapse in 2008 an increased correlation is seen between BRICS and the US [6], indicating a delayed contagion effect. This may indicate that the sensitivity of a given country's Sovereign Risk to a GFC may differ depending on its level of market development. Recent studies have indicated that the linkages between Developed and Emerging economies have been increasing over time, resulting in more correlated economic movements in response to a GFC. A study conducted by Edgardo et. al. investigated the response of the respective markets to a GFC and it was found that local currency denominated Sovereign Bond yield spreads of Emerging Markets were least affected between crisis and non-crisis periods. However in a GFC the exchange rate of Emerging Markets typically plummet, with money being pulled out of the country and rather placed in more stable dollar-denominated products. The foreign holders of locally denominated bonds experience a large capital loss as a result of the local currency exchange rate collapse.

It can be seen that the level of inter-connectivity between markets is increasing over time. This makes Emerging Markets more vulnerable to GFCs. For example the South African rand's high liquidity, resulted in it plummeting severely in the 2008 financial crisis resulting in a large increase in South African Sovereign Risk.

1.2.2 | Sovereign & Yankee Bonds

A Sovereign Bond is a debt security that is issued by a country's government. It follows that Sovereign Bonds can thus be used to measure Sovereign Risk. This security can be denominated in either its domestic currency or in a foreign currency.

For countries with an unstable economy a bond is typically issued in a foreign more stable currency that has no direct exposure to domestic inflation risk, rather than domestic currency where there is direct exposure. However if a country's currency depreciates it makes it more difficult to service offshore debt unless hedged with offshore income. For the purpose of comparison, this research considers only USD denominated Sovereign Bonds. This provides a common base currency, the USD, across which to compare Sovereign Risk. All bond issuers face exchange rate risk relative to the USD and all bond holders enjoy the stability of receiving loan repayments in USDs.

The default risk of a Sovereign Bond, and hence the Sovereign Risk of that Bond, can be measured through the yield of that bond. The yield being the interest paid to an investor that has purchased a bond. The riskier the bond is perceived

to be the higher the required rate of return (yield) of that bond. Given the inverse relationship between the yield and price of a bond, riskier bonds are typically priced at a discount to par.

A Yankee Bond is also a foreign bond that is denominated in U.S dollars. It is further stipulated that this bond must be issued within the U.S. and be registered with the Securities Exchange Commission (SEC). This bond can be issued by a foreign bank, foreign company or a foreign government. Given that the lowest level of granularity of this research is on a national level these bonds are not considered, only Sovereign Bonds are used in this research.

1.2.3 | Sovereign Yield Spread

In order to measure comparative Sovereign Risk, a basis of comparison needs to be determined. The U.S. Sovereign Yield is used as this base measure. The difference of each country's yield to the U.S. yield forms the yield spread measure, where one can now discern the comparative risk of each country relative to the U.S.

1.2.4 | Emerging vs Developed Economies

An economy is typically categorised into one of three groups; Emerging, Developing or Developed. This helps to give an indication of the level of development and growth potential of that economy. Emerging economies are typically those that are experiencing significant growth, while Developing nations are struggling, often requiring the aid of Developed countries [7].

The Economist describes an Emerging Market as, "an economy that is not too rich not too poor and not too closed to foreign capital." [8]. The term "Emerging Market" was coined in 1981 by Antoine van Agtmael, who was working for a division of the World Bank at the time [8]. His aim was to conveniently group countries that show promise of future growth, into a single fund. This fund would in turn encourage foreign investment. The initial fund based on this mandate, the Morgan Stanley Capital International (MSCI) Emerging Markets (EM) Index, consisted of only 4 countries. Today this number has risen to 24.

Within the parameters of the above Emerging Market definition, there is scope for significant differences between economies. The MSCI EM Index has for this reason been accused of being an "indiscriminate grab-bag," of economies at different levels of development [8]. Some differences include the size and income level of the economy as well as the growth-driving sectors of the economy.

With regard to income level, 9 countries listed on the MSCI EM Index have been classified as high-income countries according to the World Bank [8]. In isolation this could be interpreted as a Developed country characteristic. In terms of economy size, China provides an excellent example of a peculiar Emerging Market country, with its economy being the second largest in terms of nominal GDP and the largest in terms of purchasing power parity according to the International Monetary Fund [9]. Furthermore the growth drivers of an economy can differ vastly under the umbrella definition of an Emerging Market. For instance South African growth is largely driven by natural resources, while China's economy is driven by manufacturing[10].

These major differences across economies within the Emerging Market definition, allows for their behaviour to be distinctly different with regard to their Sovereign Risk. There is therefore an argument for the refinement of the types of economies within this group and their corresponding behaviours. A better understanding of these sub-categories and their corresponding risk environments can lead to more informed decisions surrounding investment as well as government strategies in obtaining foreign investment.

2 | LITERATURE REVIEW

2.1 | Yield Spread

The yield associated with a given bond acts as an indicator of the perceived risk of that bond, where the required rate of return increases with the perceived risk. A yield spread on the other hand is a measure of relative risk. The yield spread of a given country or market acts as a measure of the relative risk of that country/market to a chosen benchmark. An example of such a measure is developed by Cayon et al. (2018) based on fair market value zero coupon Sovereign Bond curves (FMCZCB), available from the Bloomberg terminal at a daily frequency. These curves are derived from real bond prices and provide bond price estimates for non-traded maturities.

The yield spread measure, applied to groups of countries, is developed by Cayon et al. (2018) as follows:

- For each group, the contribution of a country's yield to the overall group yield measure is weighted according to the country's relative economic importance within each maturity segment. The weight of country i 's Sovereign debt securities at each maturity j is calculated as follows:

$$w_{j,i} = \frac{v_{j,i}}{\sum_{j=1}^n v_{j,i}}$$

where $v_{j,i}$ is the USD denominated value of Sovereign Bonds in country i with maturity ranges $j = 1, 2, \dots, n$. (1)

and $\sum_{j=1}^n v_{j,i}$ is the sum of USD denominated Sovereign Bond values across all j maturities for a given country i .

- The value-weighted theoretical bond yield for country i , across all maturities j , at a given time t is calculated as follows:

$$Y_{i,t} = \sum_{j=1}^n w_{i,j} y_{i,j,t} \quad (2)$$

where $y_{i,j,t}$ is the Sovereign yield, corresponding to a Sovereign Bond of maturity j , for country i at time t .

It is also noted that $w_{i,j}$ is kept constant for all values of t .

- $Y_{i,t}$ can now be used to calculate country i 's daily spread at time t .

$$spread_{i,t} = Y_{i,t} - Y_{US,t} \quad (3)$$

where $Y_{US,t}$ is the United States weighted average Sovereign Bond yield.

- The yield spreads of each country, within a given group p at time t , are then weighted and aggregated as follows:

$$SPREAD_{p,t} = \sum \left(\frac{V_{i,p}}{\sum_{i=1}^p V_{i,p}} \right) spread_{i,p,t}$$

where $V_{i,p}$ is the total value of Sovereign Bonds outstanding within country i in group p . (4)

and $\sum_{i=1}^p V_{i,p}$ is the total value of Sovereign Bonds outstanding in group p .

The yield spread measure that has been developed above can now be modelled as the dependent variable against potential determinants of this spread. In the cases where the frequency of the yield spread observations and its potential determinants do not match, the more frequent measure is aggregated in order for the frequencies to match.

2.2 | Sentiment Analysis

Sentiment Analysis refers to the process of extracting the emotional content of a body of text. This is typically achieved through the development of an algorithm that systematically classifies each of the words in a body of text as positive, neutral or negative. The sentiment of each of the words can then be aggregated to form an overall sentiment score for a body of text. The algorithm can vary in complexity depending on its purpose and the desired accuracy of the sentiment score. Variations between algorithms arise in the classification categories as well as in the extent to which language rules are considered. An ordinal scale can be developed, categorising the extent to which words are positive or negative and linguistic rules such as negation can be accounted for through reversing the sentiment score of words that directly follow identified negation words.

An example of such an algorithm, named the "SentiStrength" tool, is used by Stieglitz and Dang-Xuan (2013) in order to attain the sentiment associated with politically oriented tweets. This algorithm makes use of a human designed lexicon of emotional words as well as making use of linguistic rules for negation, booster words, amplifications, emoticons, spelling corrections and word weightings [11]. Where booster words typically enhance the sentiment of the words that follow (very nice) [11] and amplifications enhance the sentiment of the given word through the manipulation of the spelling of the word (haaaaapy) [11]. The algorithm gives the body of text both a positive as well as a negative score. Each of these scores makes use of a scale. The positive score can range from 1 to 5, where 1 is a neutral score and 5 is strongly positive. Similarly the negative score ranges from -1 to -5, where -1 is neutral and -5 is strongly negative. This method of providing both a positive as well as a negative score has been proven to produce more accurate results in comparison to standard machine learning approaches [11]. Finally the overall sentiment of the tweet is established through a polarity measure developed as follows:

$$polarity = positive\ score + negative\ score \quad (5)$$

Based on this formulation the polarity measure ranges between $[-4, 4]$. This measure is however limited in recognising texts that have equally weighted positive and negative sentiments, the measure will effectively cancel out the positive

and negative sentiments to give a neutral score, when the text is in fact emotionally charged with both positive and negative sentiments [11]. Hence an additional measure was introduced by Stieglitz and Dang-Xuan (2013) as follows:

$$sentiment = (positive\ score - negative\ score) - 2 \quad (6)$$

where *positive* and *negative* represent the respective sentiment scores as before. This formulation of the measure no longer offsets the negative and positive sentiment but rather measures the extent to which the text is emotionally charged, regardless of whether the sentiment is positive or negative. Furthermore 2 has been subtracted from the measure to effectively normalise the range of sentiment scores from [2, 10] to [0, 8] since the lower bound in fact represents a neutral sentiment.

2.3 | Regression Yield Spread Modelling

Several models have been developed in order to estimate Sovereign Yield spreads [3, 12, 13]. Furthermore each study pursues a different goal and context within which to understand the behaviour of Sovereign Yield spreads. Across these models a range of potential predictors are explored, some being global and others country specific. Furthermore the yield spread behaviour across different groups of countries is also explored. A sample of these models is detailed below alongside their contributions to the literature on Sovereign Yield spread analysis.

2.3.1 | Sovereign Yield Spread over the Global Financial Crisis

A model developed by Cayon et al.(2018) looks to determine the level of financial immunity of the Emerging Markets in comparison to the Developed Markets over a period of financial turmoil. A total of nine different groups of countries are tested, to ascertain how the behaviour of Sovereign Yield spreads differs across these groups. Of the 9 groups tested in this study, the Emerging Markets, Developed Markets and USD denominated Sovereign Bonds are of interest. A panel regression model is developed and fitted to the respective groups' Sovereign Yields.

The determinants of a given group or country's Sovereign Yield spread are categorised into 3 groups. Each determinant group is named and described by Giordano et al. (2013) and Dungey and Martin (2007) respectively:

- **"Shift Contagion"** - This is a common / global shock. An example being a liquidity shock.
- **"Wake-Up Call Contagion"** - This describes a country specific shock that will effect the default risk of a particular country. Such variables include government indicators as well as macro-economic variables.
- **"Pure Contagion"** -This describes a Latent / idiosyncratic shock. This being a shock that is neither a common nor a country specific shock. It is however believed to affect all Sovereign Bond Markets simultaneously.

Using these categories Cayon et al. (2018) categorise the yield determinants tested in their study as follows:

TABLE 1 Yield spread determinants according to category

Determinant	Description
"Shift Contagion"	
Global Equity Premium	<i>S&P Global 1200 return</i> – $Y_{US,t}$. <i>S&P Global 1200</i> describes the Standard & Poors Global Index that consists of 1200 stocks. This index effectively represents the global equity market. $Y_{US,t}$ represents the value-weighted theoretical bond yield for the United States. Note that the U.S. Equity Premium is estimated by: <i>S&P 500 return</i> – $Y_{US,t}$.
Regional Bond Yield Spreads	value-weighted yield of bonds within group excluding country of interest – $Y_{US,t}$.
Global Risk Aversion	Represented by the implied volatility index (VIX) on the S&P 500 index.
Bond Market Activity	Measured through the volume of bond trading.
"Wake-Up Call Contagion"	
Local Equity Premium	Represented through the changes in the local stock market index.
Bond Market Conditions	Represented through a value-weighted average of bond price returns in each country less the U.S. equivalent measure.
Macro Economic Variables	Exchange Rate, total Debt-to-GDP, Investment-to-GDP, external Debt to Exports, GDP per capita growth, reserves and country specific bond price return.
"Pure Contagion"	
Binary Global Financial Crisis variable	Returns true when there is currently a financial crisis.

The model is then developed as follows:

$$spread_{i,t} = \alpha_i + \beta'_s spread_{i,t-j} + \beta'_l Z_{i,t} + \beta'_c F_t + \gamma_0 D_{c,t} + \gamma'_1 (Z'_{i,t} D_{c,t}) + \gamma'_2 (F'_t D_{c,t}) + \epsilon_{i,t} \quad (7)$$

where α_i is the country specific intercept. β_s a vector of coefficients on the lagged values of country i 's yield spread ($spread_{i,t-j}$), where j defines the size of the time lag. β_l and β_c are vectors of coefficients for the local ($Z_{i,t}$) and common (F_t) factors respectively. D_c acts as a financial crisis indicator variable while the γ coefficients are used to test for changes in transmission channels¹ during financial crisis periods. The γ coefficients also test for interactions between financial crises and the local and common factors respectively. The residuals, represented by $\epsilon_{i,t}$, of the model are

¹A transmission channel in this case refers to the avenue through which yield spreads are affected. Three channels are investigated here, namely the shift, wake-up call and pure contagion channels.

robust according to Longstaff (2011) and standardised.

The financial crisis variable is set to true from October 2008, in order to coincide with the failure of Lehman Brothers, to March 2013. If γ_0 is found to be significant this can be interpreted as a "pure contagion effect" [14, 15]. If γ_1 and/or γ_2 , which relate to "wake-up-call contagion" and "shift contagion" respectively, return significant results this indicates that there is a change in these factors effect on yield spread during a period of financial turbulence.

The model findings are as follows:

- Global Equity Premium proved to be significant for all 9 bond groups, with Emerging Markets and USD denominated debt yield having an inverse relationship with the Global Premium and hence with the performance of the economy.
- Amongst the country-specific factors it is found that the Exchange Rate, Debt/GDP, Current Account/GDP and local Equity Premium are significant.
- It was found that the association between macro-economic and global factors to the yield spread changed over periods of financial crisis. In the Emerging Markets it was found that the relationship between macro economic variables, namely the Exchange Rate, Investment/GDP, Current Account/GDP and the local Equity Premium, and the yield spread weakened over the GFC period. However it is noted that the spreads of local-denominated Sovereign Bonds in Emerging Markets were not significantly affected by the crisis.

The study goes on to improve the comparison between crisis and non-crisis periods by accounting for the inconsistencies that arise from the choice of crisis and non-crisis periods. The factor model is extended and a propensity score matching method is developed to select periods of crisis and non-crisis that can be compared. This however falls outside of the scope of interest for this study and is thus not elaborated further.

2.3.2 | Political Determinants of Sovereign Bond yield spreads

Eichler (2014) investigates the predictive ability of political variables in determining Sovereign Bond yield spreads. His analysis is conducted on 27 Emerging economies from 1996 to 2009. The yield spreads are calculated as the difference between the given country's Sovereign Yield and U.S. Sovereign Bond yields. The political variables that are tested can be seen below.

TABLE 2 Political variables that are considered as potential Sovereign Yield spread determinants.

Determinant	Description
Nature of Political System	A political system can range between an autocratic and a democratic regime.
Elections	Elections and whether they occur or not act as a proxy for public control over the chief executive and political business cycle ² .
Ideology	Political ideology refers to a set of beliefs related to how a country should be run. Leftist parties are typically in favour of state control over major institutions while right wing parties support the privatisation of such businesses.
Political Stability	Political Stability can refer to the length of a party's tenor as well as to the level of agreement amongst parties on how the country should be run.
Feasibility of Policy Change	This refers to the ruling party's ability to implement its policies.
Quality of Governance	This refers to the effectiveness and suitability of the policies put in place by the ruling party as well as the rights granted to the public.

Political Business Cycle ²

The motivation for including these variables ,as well as how they are measured, is elaborated below:

- **Nature of Political System:** Under a democratic regime the choice of government is delegated to the public, who then also hold the power to vote members out of power. Under an autocratic regime the public does not have this power. The Sovereign Risk may be lower under democracy due to the increased accountability of the government to the public. However an autocratic regime faces less risk in terms of political business cycles. A political business cycle refers to spending patterns that arise prior to and after government elections. Governments will typically implement an expansionary fiscal policy prior to elections in order to secure votes, resulting in the need for a contractionary policy post elections to stabilise public finances. Furthermore unpopular austerity programs are more easily implemented under an autocratic regime. Parliamentary and presidential regimes are also compared. Under a presidential regime it is easier to implement unpopular budget consolidation measures, which could decrease perceived Sovereign Risk.

The nature of a political system is measured using a **policy score**, where higher values indicate a more democratic regime and lower values an autocratic regime. The score ranges from –10 to 10 to account for political systems that lie between the two extremes.

²A Political Business Cycle is a cycle that emerges as a result of the manipulation of policy tools, such as fiscal & monetary policy. A ruling party will typically enforce expansionary policies prior to elections, to improve their chances of re-election.

- **Elections:** Political Business Cycle theory investigates the fluctuations in government revenue and spending surrounding elections. Pre-election periods often see increased government spending and reduced taxes in order to improve the chances of re-election. This behaviour is indicative of unsustainable public finances, which results in increased Sovereign Risk.

The effect of elections is measured through the use of binary variables:

- **Pre-election:** returns true (1) if the chief executive faces elections next year and zero otherwise.
 - **Legislative election year:** returns true (1) if legislative election takes place in current year and zero otherwise.
 - **Executive election year:** returns true(1) if executive elections take place in current year and zero otherwise.
- **Ideology:** The potential effect of Ideology on the financial accountability of governments is also considered.

The effect of political ideology on Sovereign Yield spreads is measured through the use of binary variables:

- **Left party government:** returns true (1) if the party is communist, socialist, social democratic or left wing. The variable will return zero otherwise.
 - **Right party government:** returns true (1) if the party is considered to be conservative, christian democratic or right wing. The variable will return zero otherwise .
- **Political Stability:** Higher political stability should be associated with lower Sovereign Risk. A government operating in a relatively stable political regime should have longer periods of time to implement its fiscal policies, resulting in lower levels of Sovereign Risk. Furthermore less frequent regime changes should incite confidence in the government's ability to maintain sustainable public finances.

The measurement of political stability considers both internal and external influences to the political system. The internal influences consist of the following variables:

- **Tenure of the government party:** This variable indicates the number of years the ruling party has been in power.
- **Checks & balances:** Describes the degree of checks and balances faced by the ruling party. The higher these are, the more challenging it is to implement new policies.
- **Veto players:** Describes the number of veto players. The more veto players present, the more challenging it is to implement policy changes that can manage the levels of public debt.

The external influences are described by the following variables:

- **Overall Political Stability Index:** This measure forms part of the *Worldwide Governance Indicators* developed by the *World Bank* and measures the probability that a government is destabilised or overthrown by violent or unconstitutional means.
 - **Internal Conflict:** This is a binary variable that returns true if there is conflict present. Examples of the types of conflict include revolutionary, ethnic or political war.
 - **Military Officer:** This is a binary variable that returns true (1) if the chief executive of the ruling party has links to the military and zero otherwise. This measures the implication for stability given that the ruling party is associated with the military.
- **Feasibility of Policy Change:** In order to incite confidence from investors the government must be able to implement necessary policy changes in order to ensure its ability to meet debt obligations. A higher government majority makes it more difficult for opposing parties to impede policy changes and will thus be associated with lower Sovereign Risk.

Furthermore the ability to come to a consensus on the most appropriate policies is easier when fractionalisation³ and polarisation⁴ within a government is low.

The variables used to measure the feasibility of policy change include the following:

- **Government Majority:** This describes the ruling party's majority in parliament and across all the houses in parliament.
 - **Government Fractionalisation:** Fractionalisation refers to the number of parties in government.
 - **Government Polarisation:** This refers to the differences in party orientation in government.
 - **Executive Constraints:** This measures the extent of institutionalised constraints, that can make it difficult for governments to implement necessary policy reforms and hence increase Sovereign Risk.
- **Quality of Governance:** Good governance can increase the sustainability of public finances through better legal processes and higher growth prospects, that can in turn reduce public Debt to GDP. Furthermore better regulations can also lead to better tax compliance and hence improve the governments' ability to meet its debt obligations.

A description of the quality of governance is formulated through the following variables:

- **Rule of Law:** Refers to the functioning of the legal system.
- **Regulatory Quality:** This variable may be associated with improved government operations that can positively influence government spending and tax collections as well as encourage economic growth.
- **Government Effectiveness:** This can refer to the efficiency of government administration, which prevents the unnecessary wastage of government funds.
- **Freedom from Corruption:** Increases in corruption can increase the cost of doing business as well as raise the levels of uncertainty. Simultaneously however it may be easier to initiate investment projects, boosting economic growth, by surpassing regulations.
- **Voice & Accountability:** This refers to the liberty rights and political participation of the public. A public that is well informed may be more supportive of contractionary fiscal policy, improving tax collection statistics.
- **Economic Freedom:** An economy that is relatively free, will have more growth opportunities compared to a highly regularised economy. However policies that are too relaxed can also hinder economic growth with rising levels of income inequality and business cycle amplitudes.

To isolate the effects of the above political determinants a number of control variables are also included in the model. The relationship between these variables and Sovereign Yield spreads has been tested in multiple previous studies and have been found to be significant [13]. Controls relevant to each country as well as global controls are accounted for.

The country-specific controls are described through the following variables:

- **External Sovereign Debt to GDP:** Higher levels of Sovereign Debt relative to a country's GDP is indicative of limited government funds to meet its debt obligations and is therefore associated with greater Sovereign Risk.
- **Short term debt to reserves:** Short term debt is included specifically to account for potential liquidity problems. Furthermore in the case where the debt is denominated in USD, one will be interested in the amount of USD held in reserve.

³Fractionalisation refers to the number of parties in government. Low fractionalisation implies few parties, while high fractionalisation indicates many.

⁴Polarisation refers to the extent to which party political views are aligned. Low polarisation indicates similar views while high polarisation suggest very different views.

- **Economic Growth & Investment to GDP:** These variables are included in order to account for the state of the given country's economy. Economies that have positive growth prospects should be able to meet their debt obligations lowering the associated Sovereign Risk of that country.
- **Openness:** This refers to the openness of the economy through international trade. The sum of total exports and imports to GDP act as a proxy for this concept. The more open the economy, the more vulnerable it is to international perception of its risk. However having access to the international market also increases the chances of being able to refinance its debt.
- **Inflation:** Higher rates of inflation result in the depreciation of the local currency and hence make it more expensive for the government to repay foreign denominated debt.

The global controls consist of the following:

- **VIX Index:** This index is an average of the implied volatility of eight put and call options written on the S&P 500 index. This acts as a measure of the expected volatility, and hence the expected risk, in the financial market.
- **TED Spread:** This is an interest spread measure between the LIBOR (interest rate on interbank loans) interest rate and U.S. treasury interest rate. It is used to gage global liquidity conditions.
- **U.S. Interest Rate:** This is used as an indicator of investors' general risk aversion.
- **Sovereign Default, Banking Crisis & Currency Crisis:** These are binary variables that are indicative of financial crisis events that are associated with increased chances of default.
- **Sovereign Default History:** A binary variable that indicates whether a Sovereign default has occurred within the previous 10 years. Including this variable explores the possibility that a default in the past, increases the chance of default in the future.
- **IMF Support:** A binary variable indicating whether a country is a recipient of IMF loans. This lending arrangement can improve a governments liquidity, however the need of a loan in order to service debt may drive up the Sovereign Yield spread.

A series of models are now formulated to test the relationship between these predictor variables and the yield spread. First a baseline regression model is developed: A fixed effects panel regression model. The formulation of this model can be seen below:

$$Spread_{i,t} = \alpha Politics_{i,t} + \sum_j \beta_j Controls_{j,i,t} + \sum_k \mu_k Controls_{k,t} + \gamma_i + \epsilon_{i,t} \quad (8)$$

where the Sovereign Yield spread of country i at time t is regressed against the political variable, $Politics_{i,t}$, and a control variable specific to country i ($Controls_{j,i,t}$) as well as a global control variable ($Controls_{k,t}$). The fixed effect related to each country is represented by γ_i . Notice that the fixed effect estimator is invariant over time and it accounts for the unobserved impact of each country on yield spread. α , β_j and μ_k are the respective coefficients of the predictor variables that need to be estimated and the residuals of the fitted model are represented by $\epsilon_{i,t}$. The t-values are based on standardised and autocorrelation-robust errors.

Furthermore it is verified that a fixed effect model is the appropriate model to be using, as a random effects model is also fitted to the data and it was found to produce inconsistent effects according to the Hausman (1978) test. Five more sensitivity tests are conducted, to assess the robusticity of the results and it is concluded that the findings obtained in the baseline specifications are robust. Each test makes modifications to the model (Equation 8) as specified above and takes note of the resulting change. Details of the five tests are as follows:

1. The mean value of the daily EMBI spreads is used as the dependent variable, as opposed to the end of year value used in the original model.
2. Additional control variables are introduced:
 - Debt service to GDP ratio: public and publicly guaranteed debt service, which is made up of interest and principal payments.
 - External financial requirements ratio: short term Sovereign debt minus current account balance to foreign exchange reserves (proposed by Manasse and Roubini (2009)) is used as opposed to the current account balance.
 - Percentage devaluation of domestic currency against the USD is used as opposed to the inflation rate.
 - Rates of return on Merrill Lynch High Yield Bond Index is used instead of the VIX index.
3. GDP growth projected one-year ahead using constant prices is used instead of actual GDP growth.
4. Observations falling within one year after a Sovereign default are excluded from the model.
5. Each explanatory variable is lagged one year behind the dependent yield spread variable.

The findings of this model are displayed in the tables below:

TABLE 3 Findings with regard to political variables

Determinant	Finding
Nature of Political System	Non-significant, however within a democracy, a significant difference arose between a parliamentary and presidential system, with parliamentary systems being associated with higher yield spreads.
Elections	Non-significant
Ideology	Significant, higher Sovereign Yield spreads are associated with left and right party governments in comparison to centre governments.
Political Stability	Significant, increased stability is associated with lower yield spreads.
Feasibility of Policy Change	Non-significant
Quality of Governance	Significant, higher quality is associated with lower Sovereign default risk. However no significant association is found between corruption and Sovereign Yield spreads. Furthermore Economic Freedom is also found to be non significant.

TABLE 4 Findings with regard to country-specific control variables

Determinant	Finding
External Sovereign Debt to GDP	Significant, higher levels of external Sovereign Debt to GDP are associated with larger yield spreads.
Short term debt to reserves	Non-significant, this indicates that overall indebtedness affects Sovereign Risk but not the maturity structure of the debt.
GDP growth & investment to GDP	Non-significant and significant, indicating that the prospects of future growth effect the Sovereign Risk while current growth does not.
Openness	Non-significant
Foreign exchange reserve to imports & current account balance	Both significant, with higher levels of both measures associated with lower yield spreads.
Inflation rate	Non-significant, indicating that investors do not condone seignorage as a sustainable means of maintaining Sovereign solvency. Furthermore in the case of foreign denominated debt, an increase in money supply is often followed by a depreciation in domestic currency making it more difficult to repay foreign denominated debt.

seignorage⁵

TABLE 5 Findings with regard to global control variables.

Determinant	Finding
TED spread & U.S. interest rate	Significant, a higher TED spread and a lower U.S. interest rate is associated with increased Sovereign Risk
VIX Index	Non-significant
Sovereign default dummy	Significant, with Sovereign spreads approximately 13% higher in comparison to non-default periods
Default history	Non-significant
banking & currency crisis	Non-significant
IMF program	Non-significant

Tests for a non-linear relationship between the Sovereign Yield spread and the chosen predictors are also conducted. By applying the natural log transformation to the dependent variable, increasing returns of the respective political variables are tested. Having applied this transformation it was found that government fractionalisation and polarisation

⁵ seignorage refers to the profit made by the government through issuing more currency.

become significant in determining Sovereign Yield spreads. This indicates that only extreme levels of fractionalisation and polarisation will have an effect on Sovereign Risk. Similarly it was found that corruption and economic freedom also experience increasing returns and are thus only significant for extreme values.

The model is now modified to account for conditional affects as it now accounts for interactions between variables as follows:

$$Spread_{it} = \alpha_1 Politics_{i,t} + \alpha_2 Regime_{i,t} + \alpha_3 Politics_{i,t} \times Regime_{i,t} + \sum_j \beta_j Controls_{j,i,t} + \sum_k \mu_k Controls_{k,t} + \gamma_i + \epsilon_{i,t} \quad (9)$$

The model now considers the fact that the political variables' effect on Sovereign Risk may be dependent on the type of political regime and the degree of economic openness. The $Regime_{i,t}$ binary variable is now included, which will indicate an autocratic regime if the *polycyscore* lies between -10 and 5 and democratic if the score lies above 5 . This variation of the model produced the following results:

TABLE 6 Findings of interaction regression model

Determinant	Finding
Political Stability	Significant in an autocratic regime while non-significant in a democratic regime.
Feasibility of Policy change	Significant in an autocratic regime while non-significant in a democratic regime.

When considering the openness of the economy the model is modified; the $Regime_{i,t}$ predictor is now used as an indicator of the openness of the economy. Openness is defined as the ratio of exports plus imports to GDP. An openness ratio above the sample median is classified as open and values below the median as closed.

This variation of the interaction model indicated that the more open the economy, the less influence political variables have on the Sovereign Yield spread. It has been found that in an open economy Sovereign default may be associated with the loss of foreign trade. Consequently the ruling party actively tries to prevent default and political variables are thus less significant in determining Sovereign Risk.

3 | DATA COLLECTION & EXPLORATION

The data collected in this study is on a country level. The chosen countries are categorised as Emerging or Developed; allowing for aggregate measures to be formulated for both the Emerging and Developed Markets. Depending on the frequency of the data collected, the data is aggregated to a monthly frequency for short-term analysis and to a yearly frequency for longer term analysis. In some cases the data is only recorded on a yearly basis, limiting its usefulness to long-term insights only.

In order to address the research question, of determining and comparing Sovereign Risk, information on Sovereign Yields is collected. This yield is then used to calculate a Yield Spread measure which is used as a proxy for Sovereign Risk. Furthermore potential Political and Economic predictors of this Yield Spread is also collected. Further details relating to the chosen countries, independent and dependent variables is discussed below.

3.1 | Country Selection

The Emerging and Developed Markets used in this study are chosen using a stratified sampling method. The resulting country choices as well as their geographical positions can be seen below.

TABLE 7 Summary of Emerging and Developed Market countries used in this research

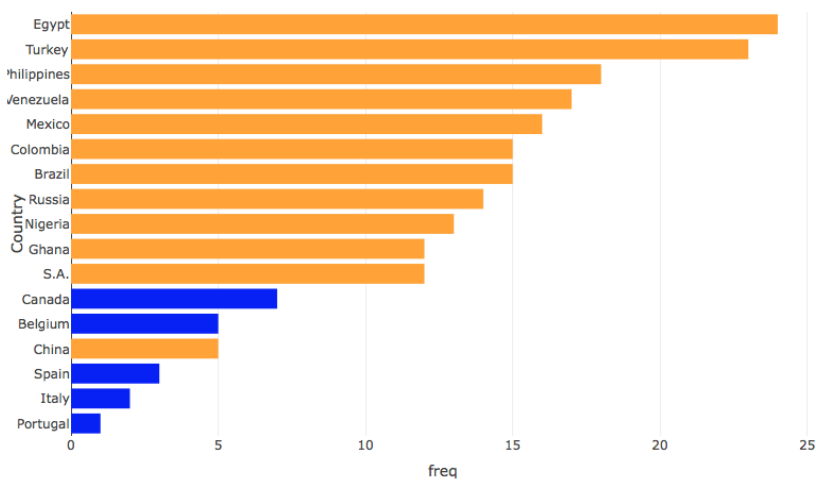
Emerging Markets	Developed Markets
Brazil	Belgium
China	Canada
Colombia	Italy
Egypt	Portugal
Ghana	Spain
Mexico	United States
Nigeria	
Philippines	
Russia	
South Africa	
Turkey	
Venezuela	

FIGURE 1 Map pinpointing sampled countries across the globe.



Countries are selected to represent each major geographic region. This sampling strategy enables one to test whether the geographic region of the respective markets has any influence on the Sovereign Risk. Furthermore the sample sizes of bond collections from the respective countries is depicted below. The maturity terms of these bonds is also represented graphically.

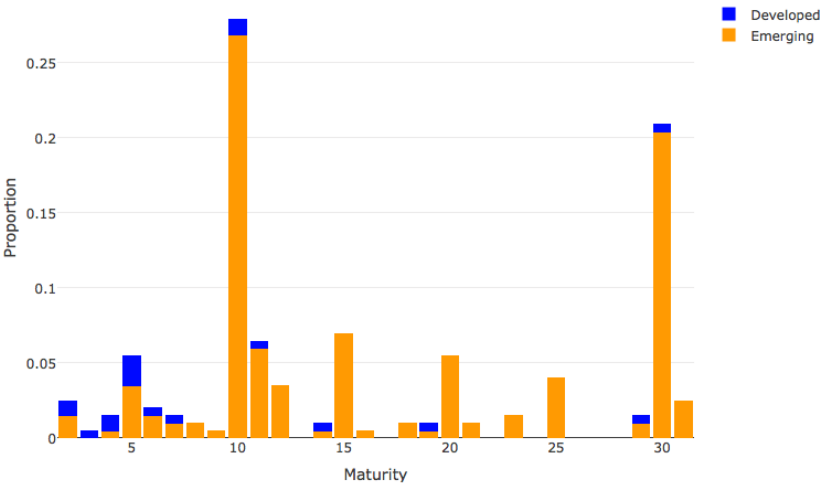
FIGURE 2 Number of active Sovereign Bonds across different countries as of August 2018.



It can be seen that there are far fewer active USD denominated bonds in the Developed Markets. Developed Markets have much larger debt stockpiles in comparison to Emerging Markets as indicated by the Global Debt Database made available by the IMF. This debt is typically in local currency explaining the smaller number of USD denominated bonds for these countries.

It can be seen that bond maturities of 10 and 30 years are most common in the Emerging Markets (*figure 3*), whereas in the Developed Markets the lending period tends to be shorter, with maturities of 5 and 10 occurring most often. The U.S. benchmark for this research is not included in figure 2. A non-exhaustive sample of U.S. bonds was collected amounting to a total of 100 bonds. Within this sample there is a variety of maturity lengths.

FIGURE 3 Sovereign Bond maturity distribution across different markets for active Sovereign Bonds as of August 2018.



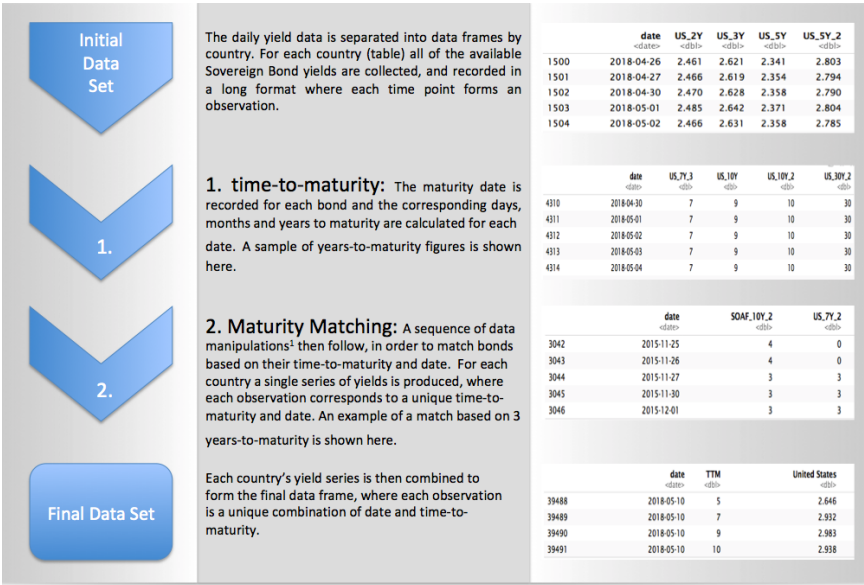
3.2 | Yield Data

Yields associated with U.S. denominated Sovereign Bonds are collected for this study. The yield time-series are sourced from the Bloomberg terminal at a daily frequency. Mid yield-to-maturity (Mid YTM) values have been selected, as apposed to the bid or ask yields. This yield acts as an estimate for the realised yield. Realised yield is of interest since it is the most accurate reflection of the default risk associated with the debt issue. A bond holder will typically drive the yield upward, while the issuer will wish to keep the yield as low as possible. The realised yield is the resulting compromise the two parties reach that should accurately reflect the default risk of the bond. YTM describes the estimated return of a bond that is held to maturity, expressed as an annual rate. Furthermore YTM is the most appropriate measure of yield for this study, since it allows one to compare bonds with different coupon rates. This increases the sample size of bonds that can be compared. For each country considered all active Sovereign Bond data is collected, as of August 2018.

In order to accurately compare the country or market specific Sovereign Risk, other factors affecting yield must be accounted for. Specifically the time-to-maturity of the bond needs to be considered. Typically the longer the time-to-maturity the riskier the bond is perceived to be and the higher the resulting yield value. By ensuring that the time-to-maturity of the bonds is the same across comparisons one is able to avoid this confounding effect.

Below is a flow diagram that describes how the yield data has been transformed in order to ensure that yields can be compared across the same time-to-maturity at a given date.

FIGURE 4 Yield Transformation Flow Chart.



¹ An R package, *YieldAnalysis*, has been developed that contains all of the functions used for the above data manipulations. A more in depth description of the sequence of data transformations can be viewed in Appendix A.

3.3 | Yield Determinants

3.3.1 | Economic & Political Data

Economic and Political variables are considered as potential predictors of Sovereign Yield Spread. The *Literature Review* conducted above is used to guide the selection of both economic and political variables, the aim being to leverage these past results. The economic variables tested aim to relate the economic health of a country to its Sovereign Risk and hence its likelihood of default. Given the insight gained from the literature review as well as economic circumstances particular to South Africa, the following economic variables are considered in this study;

TABLE 8 Economic Predictors of Interest

Yield Determinant	
Monthly	Yearly
Global Equity Premium	International Debt Issues to GDP(%)
Local Equity Premium	Current Account balance (% of GDP)
Nominal Exchange Rate	Foreign Direct Investment, net inflows (% of GDP)
TED spread	Tax Revenue (% of GDP)
	Taxes on International Trade (% of revenue)
	Total Reserves includes Gold (% of GDP)
	Total Reserves (% of total external debt)

The determinants above have been collected on either a monthly or a yearly frequency. Short term affects are analysed on a monthly basis, with measures such as the daily nominal Exchange Rate, Market indices and the TED Spread being aggregated to a monthly basis. While Long-term analysis is conducted on a yearly basis. The yield data is also aggregated, to a monthly and yearly basis respectively, in order to match the frequency of the predictors in the models that follow in this study. In previous academic work [12] it was found that Global Equity Premium, Local Equity Premium, the nominal Exchange Rate, Debt/GDP, Current Account balance /GDP and the TED Spread significantly affect Yield Spreads and hence these variables are considered in this study. Furthermore it was found that the overall debt levels as well as the flow of funds into and out of a country relative to the GDP is significant. However further investigation is conducted in the possible affects of the level of foreign investment, tax policies and the level of reserves is considered. These avenues of interest are explored through the following variables; Foreign Direct Investment, net inflows (% of GDP), Tax Revenue (% of GDP), Taxes on International Trade (% of revenue) ,Total Reserves includes Gold (% of GDP) and Total Reserves (% of total external debt) .

South Africa tax levels are relatively high in comparison to other countries [16] and given that South Africa is currently trying to secure foreign investment in the country, with President Ramaphosa scouting for investment from China. The significance of the above variables should provide for interesting insight into the risk of the South African market.

The nominal Exchange Rate is considered, as opposed to the real effective rate. Nominal and Effective rates both fluctuate in the short-term, providing an equally sensitive measure of economic sentiment at a point in time. However the real exchange rate is not available at a daily frequency. Similarly the market reacts to changes in market conditions immediately, allowing one to assess a potential relationship between the market and the associated Sovereign Risk.

The political variables considered are as follows;

TABLE 9 Political Predictors of Interest

Yield Determinant
Effective Governance
Regulatory Quality
Rule of Law
Political Stability
Voice & Accountability
Political Sentiment Score

Political measures found to be significant by Eichler(2014) are included. These variables are; The Quality of Governance, [13], as developed by the World Bank, with scores relating to the Control of Corruption, Government Effectiveness, Political Stability and Absence of Violence, Regulatory Quality, Rule of Law and Voice & Accountability.

This study furthermore makes use of Sentiment Analysis techniques to provide more insight into the political climate. The sampling process of this data can be seen under: *Poltical News Data* below. A Political Sentiment Score is developed and tested as a potential predictor of Sovereign Yield Spread.

A summary of the economic and political variables used in this study alongside their origin can be found in appendix B.

3.3.2 | Political News Data

In order to assess the impact of political news on the respective Sovereign Markets and specifically the South African market, political news articles have been collected from the *The Economist*. News falling under the sub-headings;*United States, The Americas, Asia, China, Middle East and Africa and Europe* has been collected. These stories are then filtered to only include the respective countries, and hence markets, of interest. The filtering method is based on the country name being mentioned within an article. It is therefore important to note that one article can contribute to multiple countries sentiment scores if the article mentions multiple countries. The process of scraping and filtering the news is conducted through the use of the *SentimentAnalysis* R-package that has been developed for this research. Information with regard to the article date, title and body of text is collected dating from 2007 to 2018. Regular expressions are used to clean the text and format the data appropriately.

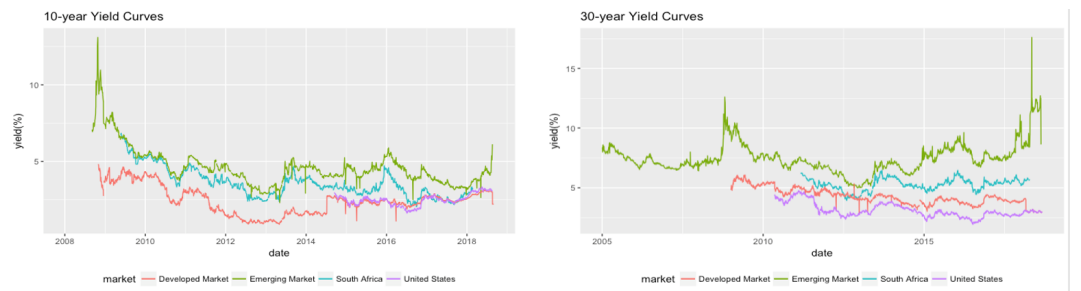
3.4 | Exploratory Analysis

3.4.1 | Yield Data

The yield data is initially explored through simply plotting selected bond yield curves. Both 10 and 30 year yield curves for the respective countries and markets are plotted below. It can be seen that relative to the Emerging Markets as a whole South Africa generally experiences lower yields, with its yield curve below that of the Emerging Markets

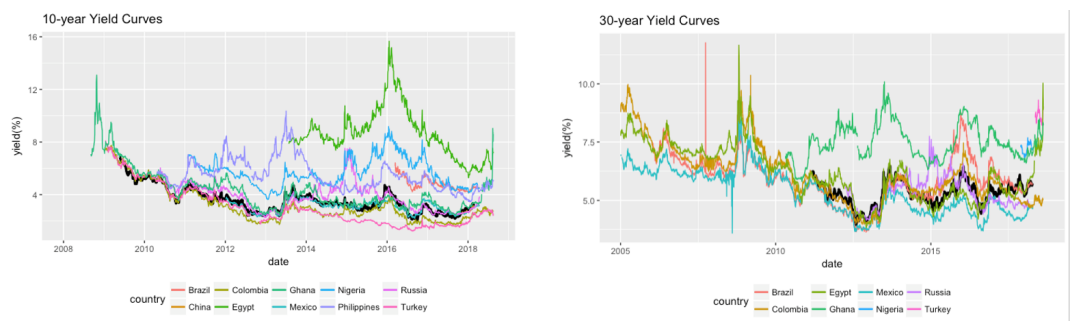
aggregate for both 10 and 30 year maturity yield curves. South Africa and the Emerging Markets aggregate experience greater yields in comparison to the Developed Markets aggregate and the United States. It can also be seen that the US yield curve acts as lower bound to the Developed Markets yield in most instances. Furthermore the United States yield is effectively a lower bound for the yield curves in the Emerging Markets as well, allowing one to use this yield as a benchmark. Analysing the overall trend of the yield curves over time, one can see that there is a significant spike in the yields over the 2008-2009 period, this is in line with the financial crisis that was experienced during this time.

FIGURE 5 South African Position & Comparison to the Emerging & Developed Markets for the period 2009 - 2018 & 2005 - 2018 for 10 year and 30 year bonds respectively.



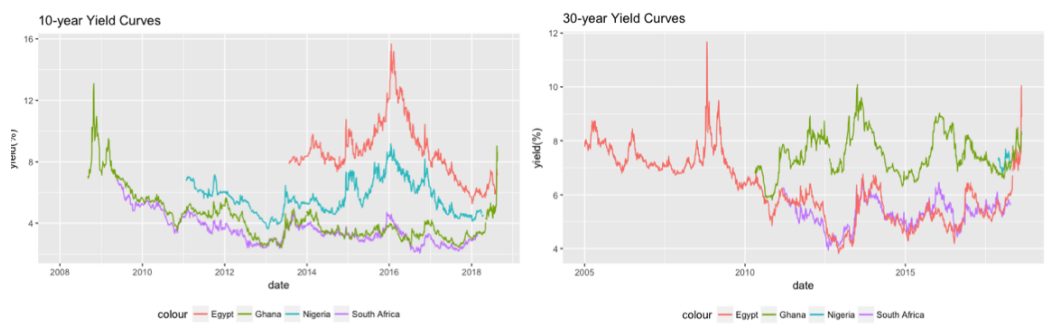
Further insight into the position of the South African Sovereign Bond Market, depicted in black, within the Emerging Markets is explored below (figure 6). South Africa is also compared to other African countries (figure 7) as well as to Mexico.(figure 8). Effectively giving insight into geographical and economic similarities on the resulting yield experienced in the respective countries.

FIGURE 6 South African Sovereign Risk Position within the Emerging Markets for the period 2008 - 2018 & 2005 - 2018 for 10 year and 30 year bonds respectively.



From the above graph it can be seen that South African yields tend to the lower bound of the group of yield curves, with Mexico being the exception.

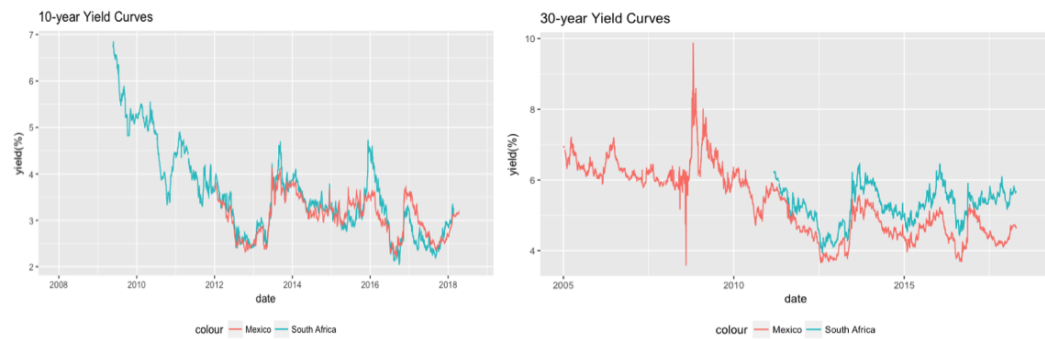
FIGURE 7 South African Position within Africa for the period 2008 - 2018 & 2005 - 2018 for 10 year and 30 year bonds respectively.



In this graph it becomes clear that the risk environment in South Africa and Egypt is similar over the longer term, as their 30-year yield curves are closely intertwined. However the Sovereign Risk is perceived to be significantly different over a shorter term of 10 years, with South Africa having significantly lower yields in comparison to Egypt for this bond maturity length. Conversely it can be seen that Ghana has a similar risk environment to South Africa in the short run (maturity of 10 years), but is perceived to be riskier than South Africa in the long run (maturity of 30 years). Nigerian risk is persistently higher in comparison to South Africa, regardless of the time to maturity.

Given the similarities of the South African and Mexican economy, the comparison of their Sovereign Risk is of interest. Both economies are resource based, have a large disparity in income between the rich and the poor and a highly volatile exchange rate [17, 18, 19, 20]. It can be seen below that the perceived Sovereign Risk of South Africa and Mexico is very similar in both the short and long run.

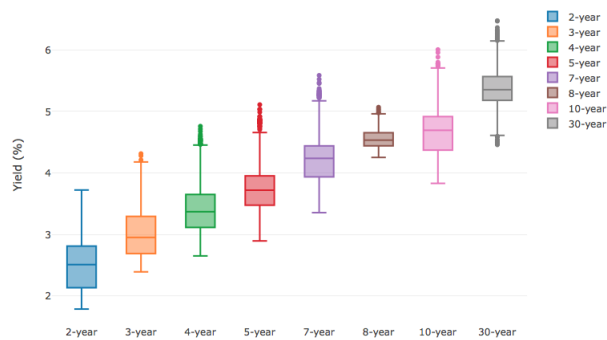
FIGURE 8 South African vs Mexican Sovereign Risk for the period 2008 - 2018 & 2005 - 2018 for 10 year and 30 year bonds respectively.



The behaviour of South African yield curves across different maturities is displayed below, as well as the coefficient of variation associated with these yield curves. From the box plot below it can clearly be seen that as the time-to-maturity

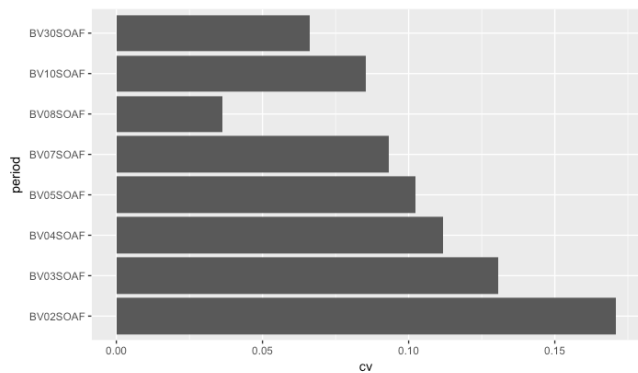
increases, the associated risk increases and hence the yield of the bond also rises. This solidifies the need to account for the time-to-maturity when modelling yield curves.

FIGURE 9 South African Sovereign Bond Maturities as indicated by SA Sovereign Yield curves for the period 2014 - 2018.



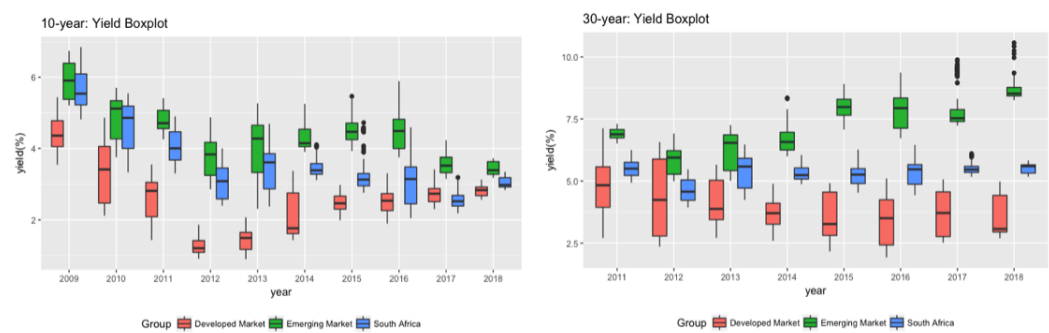
The plot of coefficient of variation below, gives an indication of the variability of yield curves with differing times to maturity.

FIGURE 10 Yield Coefficient of Variation for respective South African Sovereign Bond Maturities as indicated by SA Sovereign Yield curves for the period 2014 - 2018.



From the above two figures (*figure 9* and *figure 10*) it is noted that the yield curve for 8 years-to-maturity does not follow the overall trend that is clearly depicted by the other yield curves. This anomaly is a result of faulty data.

FIGURE 11 Yield summary for the respective markets for the period 2009 - 2018 & 2011 - 2018 for 10 and 30 year bonds respectively.



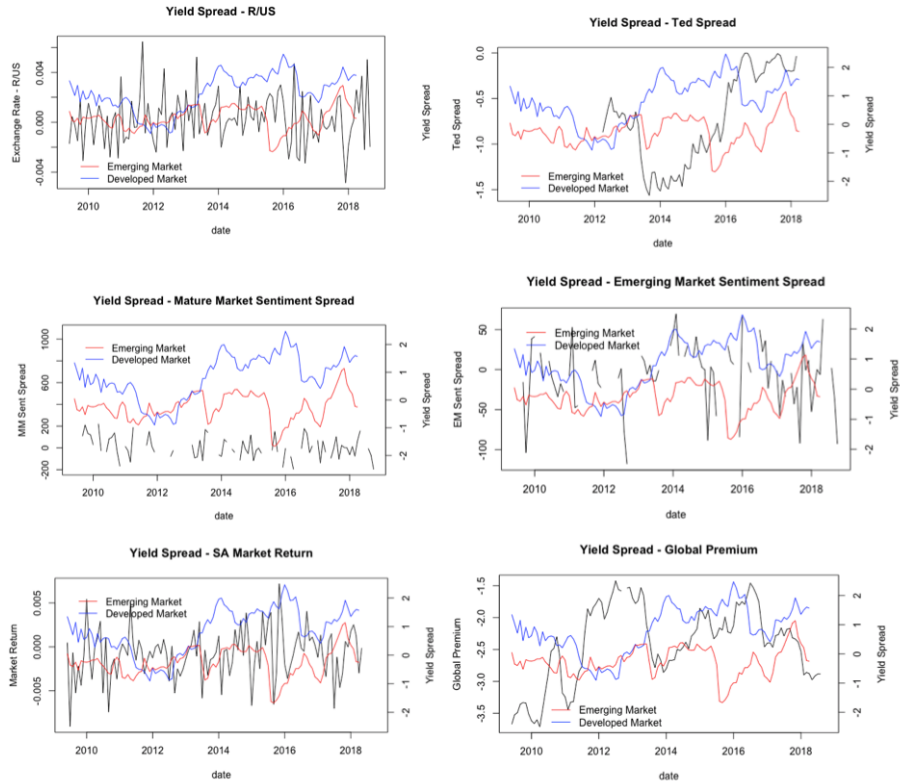
The box plots in figure 11 confirm the general ranking of yield values that is established from the yield curves of the selected bonds above; that South Africa tends to the lower bound of the Emerging Market yields and that both South Africa and the rest of the Emerging Markets experience larger yields in comparison to the Developed Markets. For the 30-year yield data an increasing divide between the yield values can be seen over time, with the Emerging Markets yield rising, and becoming relatively more risky, and the Developed Markets yield falling, becoming relatively less risky. Over the shorter term the trend seems to be more in unison. It is also noted that the variation of the Developed Markets yield is particularly large in comparison to the Emerging Markets. This may be the result of the inclusion of countries such as Italy and Spain, which are experiencing financial turmoil brought on by the financial crisis of 2008-2009.

All analysis above has had the Venezuela yield curve removed, since it is not characteristic of Emerging Markets and effectively skews the yield levels upwards for this market. Thus it will not be used in the modelling going forward in this paper.

3.5 | Association Visuals

The yield spread measures are plotted against each of the covariates, in figures 13 and 14 below, to ascertain whether an association exists between the variables. It can be seen for the short term case that the variation of the predictors is quite erratic, making it difficult to visually see trends forming between the predictors and yield spread measures. However it is noted that the Market Return variable seems to track the yield spread measures fairly well. The variation in the Market Return increases from 2014 onwards alongside an increase in the difference between the two yield spread measures. Investigating the Emerging Markets Sentiment Score one can also see some of the peaks and troughs of this series aligning with that of the yield spread measures.

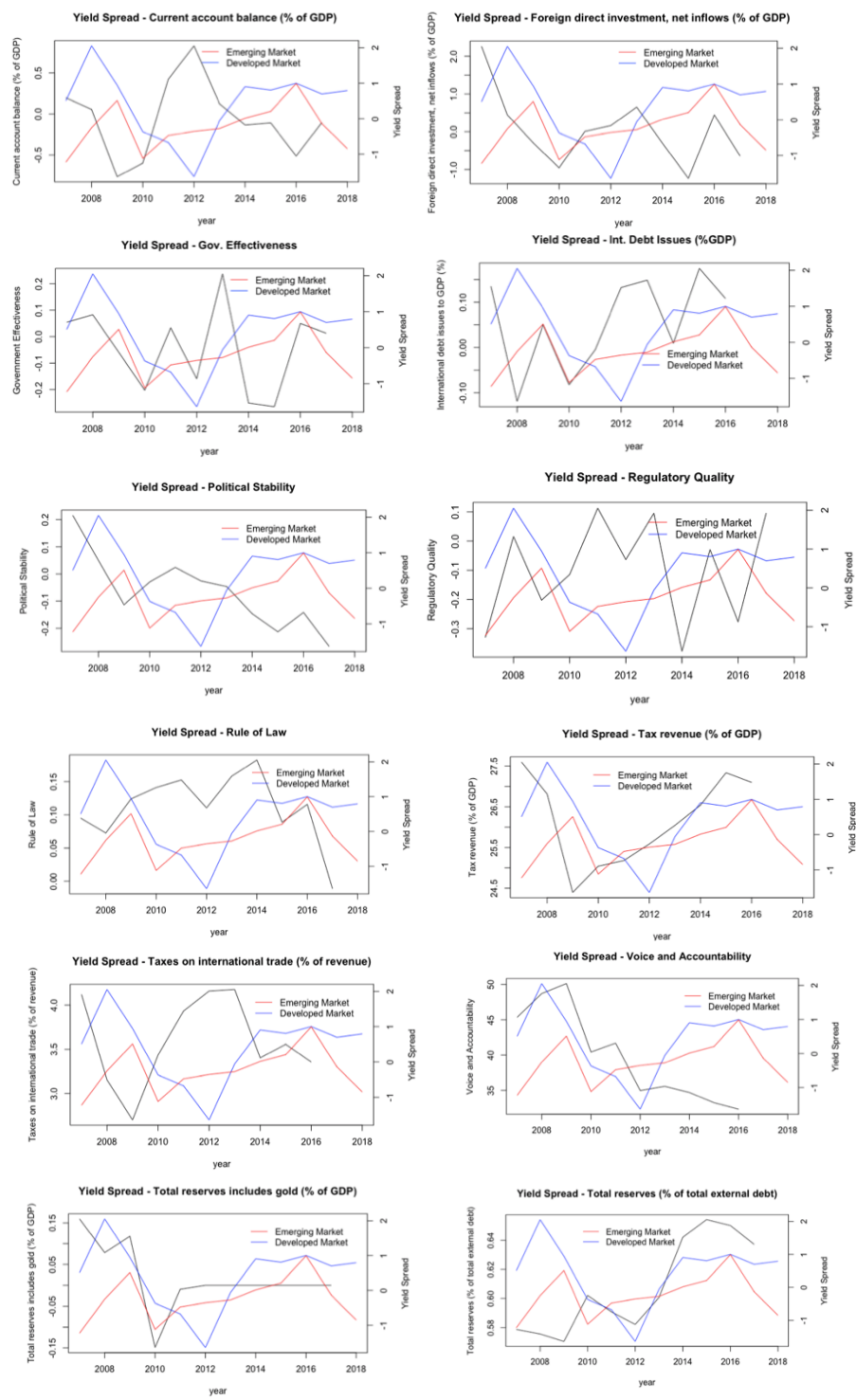
FIGURE 12 Association plots for the respective short term predictor variables.



For the long term case first note that the yield spread measures follow a similar pattern as to what was observed for the short-term case, with the series crossing over between 2011 and 2012. Indicating that during this period the difference between South Africa and the Emerging Markets was in fact larger than its difference to the Developed Markets.

A positive association between Tax revenue (% GDP) and Reserves (% of external debt) with the yield spread differences is evident. The other variables show indications of being potential predictors but the variation is quite erratic making it difficult to ascertain the relationship visually.

FIGURE 13 Association plots for the respective long term predictor variables.



3.5.1 | News Data

The distribution of the number of articles across the different countries can be seen below. All articles are collected from the Economist. Only one source is used in order to provide a single base over which to compare the sentiment across different markets. China and Russia are reported on significantly with their reports accounting for a large portion of the total number of articles collected. Furthermore the number of articles published for the period 2007 - 2018 is also depicted below. It can be seen that the Emerging Markets are reported on more regularly, providing comparatively more data on which to develop a sentiment score. Furthermore there is a spike in the number of articles collected over the period 2011 and 2012, with the number of stories relating to the Emerging Markets rising more than those relating to the Developed Markets. Furthermore multiple references to one country at a point in time have been removed.

FIGURE 14 Article Collection by Country for the period 2007 - 2018

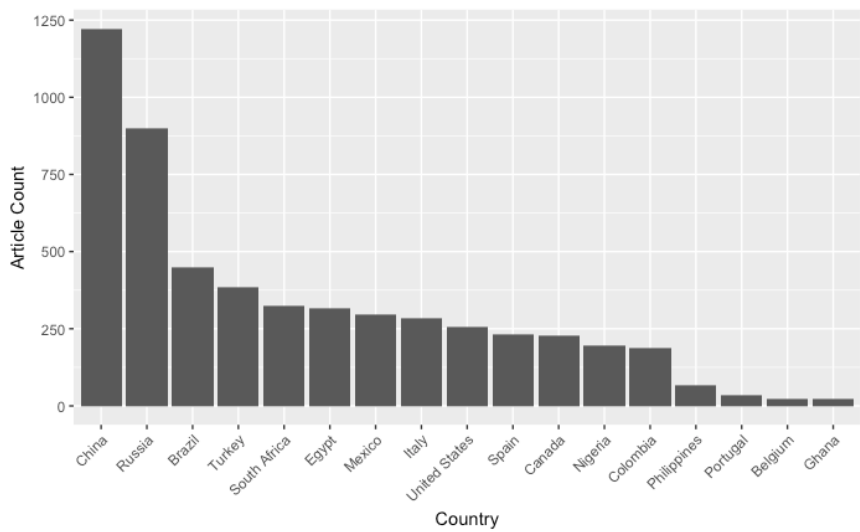
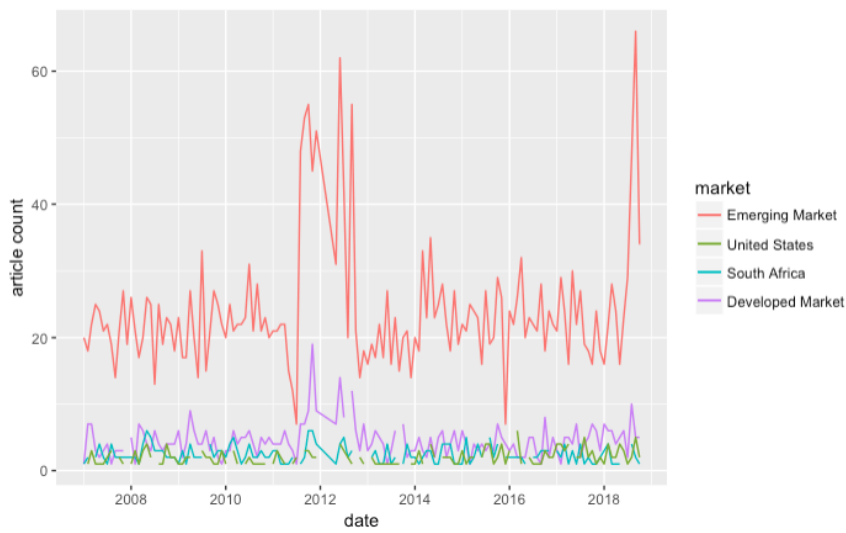


FIGURE 15 Article Collection for the period 2007 - 2018



4 | METHODOLOGY

Effectively two models are developed in this research. The first compares South African Sovereign Risk to other Emerging Markets while the second compares South African Sovereign Risk to the Developed Markets. These two relationships are evaluated on a yearly basis for the long term case of a 30 year bond and on a monthly basis for the short term case of a 10 year bond. Depending on the granularity of the analysis, predictor variables with corresponding granularity are considered. The yield spread measure as well as the sentiment score development is described below. Subsequent to this the modelling process consisting of Principal Component Analysis (PCA) and regression alongside multiple optimisation techniques is explored.

4.1 | Yield Spread Measure

The yield spread measure developed below incorporates concepts explored by Cayon et al. (2018). However in this case particular interest is taken in short and long term bond yields. For the short term case only bonds with a maturity of approximately 10 years are considered and similarly for the long term case bonds of approximately 30 years to maturity are considered. The South African yield for the short and long term case is then computed as an average of the short and long term yields at each point in time respectively (Y_{10} and Y_{30}). Similarly a long and short term yield measure is developed for the United States. Given the interest in comparison to the Emerging and Developed Markets, a yield measure is developed for these markets as well, where each country's short and long term yield values are averaged for each point in time respectively.

These yield measures, for South Africa and the respective markets, are now used to calculate yield spreads relative to the United States, as follows:

$$Spread_{i,t} = Y_{i,t} - Y_{US,t}$$

(10)

where $Y_{US,t}$ is the yield measure for the United States
and i represents either SA or the respective markets of interest
at a given time point t .

These risk measures are now further differenced in order to assess the South African risk environment relative to the Emerging and Developed Markets at each time point t as follows:

$$Market\ Yield\ Difference : Y_{SA,t} - Y_{i,t}$$

(11)

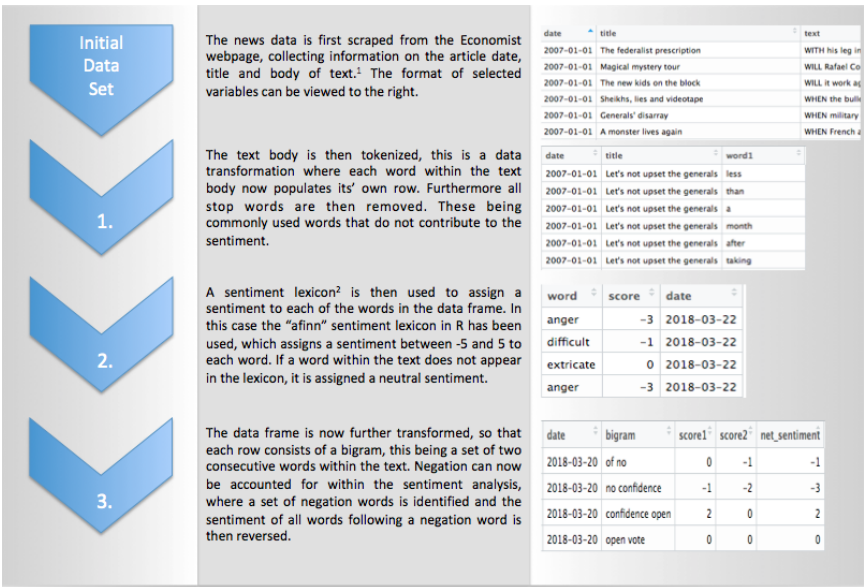
where $i = \{\text{Emerging Markets, Developed Markets}\}$

Furthermore these spread measures can then be aggregated to a monthly and yearly spread measure by averaging the daily yields across the month or year period.

4.2 | Sentiment Score Development

The sentiment extraction process is summarised in the flow chart below. This provides an over-view of the data manipulations required to calculate the sentiment scores of each pair of consecutive words scraped from the *Economist* articles.

FIGURE 16 Summary of the Sentiment Extraction Process.



¹An R package, *SentimentAnalysis*, has been developed that contains all of the functions used to scrape the news data.

²A lexicon refers to a dictionary of words, each with a sentiment score attached to it.

The bigrams as described in the flowchart above are used in order to calculate the sentiment scores across time. The use of bigrams as opposed to unigrams is superior due to the ability to account for negation, which in turn results in more accurate sentiment scores.

Using the "afinn" lexicon, as indicated above, each word within the bigram is given a sentiment. Furthermore the sentiments of words directly following negation words are reversed. The set of negation words identified are: { not, no, never, without }. Given that a sentiment has been attached to each word within the bigrams, the net sentiment for each bigram is calculated as follows:

$$Sentiment = Word\ 1\ Sentiment + Word\ 2\ Sentiment \tag{12}$$

The sentiment for each day is then calculated by aggregating the sentiments of all the bigrams that correspond to the

chosen date. These values are then further aggregated to represent the sentiment associated with a given month. Note that the sentiment scores are normalised based on the number of articles collected within the given month. An exhaustive sample was taken from *The Economist* editions and since only the number of articles collected are accounted for, the length of the articles will contribute to the severity of the sentiment score.

$$\text{Monthly Sentiment} = \frac{\sum_{j=1}^n \text{NetSentiment}_{i,j}}{\text{Total article count}_j} \quad (13)$$

where n signifies the number of bigrams in the given month.
and $j = \{\text{South Africa, Emerging Markets, Developed Markets}\}$

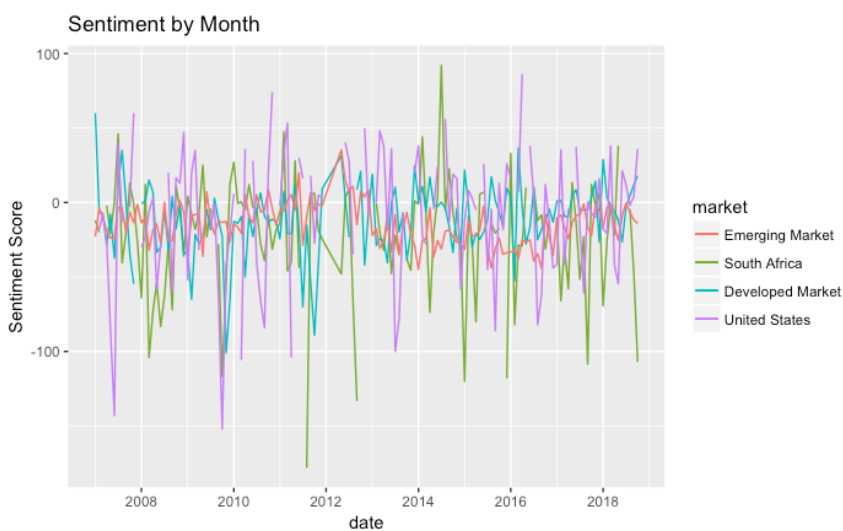
In order to compare the sentiment of South African news in comparison to the Emerging and Developed Markets respectively the collected articles from *The Economist* are first filtered in order to contain only articles relating to South Africa or the respective markets. The sentiment is then calculated for each of these groups as described above. The articles are sorted based on the article containing key words. South African articles are identified through South Africa being mentioned at least 4 times within the body of text. Furthermore if an article mentions South Africa fewer than 4 times but contained the following words: *Apartheid, Zuma, Ramaphosa, Mbeki, Mandela, Botha, de Klerk, Gupta, Gupta's, state capture, Gordan, Motlanthe* then it is also included in the South African library of articles. These extra words have been added for their political affiliation and to ensure that the South African sentiment is accurate going forward. The Emerging Market and Developed Market is measured over multiple countries hence extra search terms are not necessary. If South Africa is mentioned in the title then it is automatically included in the library. The respective country libraries are built in a similar fashion, however the matching process is simplified since a match is made based on the country being mentioned at least 4 times or if the country is mentioned in the title. The libraries for each country are then aggregated by market. Since the sentiment score is normalised by the number of articles collected, the market scores can be compared to South Africa's score.

An extract of South Africa's sentiment scores relative to the groups of interest can be seen below. The sentiment scores have been aggregated by month, since this produces a more stable trend in comparison to daily sentiment scores. Given that the number of articles collected for Emerging Markets is comparatively higher, the sample over which the score is calculated is larger, resulting in a more accurate measure. It can also be seen that the volatility of the South African and United States sentiment is significantly larger than the overall markets, see table 10 for standard deviation measures. This is intuitive since a group of countries makes it possible to cancel out each others local risk. Notice that from 2012 onwards the Sentiment Score for the Developed Markets can be seen to be above that of the Emerging Markets. The sentiment in the Emerging Markets is more pessimistic on average than in the developed markets, as indicated by the mean sentiment scores in Table 10.

TABLE 10 Summary Statistics for the respective sentiment series

series	Min	1st Q	Median	Mean	3rd Q	Max	Standard Deviation
South Africa	-310.70	-40.00	-18.5	-25.90	0.00	92.00	50.06
United States	-152.00	-31.00	-3.00	-7.69	20.67	86.00	46.24
Emerging Market	-51.00	-28.44	-16.34	-16.92	-6.79	40.00	16.454
Developed Market	-138.78	-27.94	-8.00	-12.91	3.79	109.50	32.17

FIGURE 17 Political Sentiment for the period 2007 - 2018 for South Africa, the United States and the respective markets of interest.



4.3 | Principal Component Analysis (PCA)

PCA is a method used to summarise the information contained within a large set of correlated variables. Essentially this set of variables is summarised through a smaller number of variables that encapsulate most of the variation in the data. The principal components of a dataset can therefore be understood to be the directions in the feature space along which the original data is highly variable. This is an unsupervised approach since it only considers the set of features within the data set: X_1, X_2, \dots, X_p and not the dependent variable Y . Furthermore PCA provides a means for which to visualise highly dimensional data: through extracting the directions that explain most of the variability in the data, these

directions can then be plotted in a lower dimensional space.

Each of the dimensions found through PCA is a linear combination of the original p features. The first principal component is a normalised ($\sum_{j=1}^p \phi_{j1}^2 = 1$) linear combination of the features as follows:

$$Z_1 = \phi_{11}X_1 + \phi_{21}X_2 + \dots + \phi_{p1}X_p \quad (14)$$

where the above coefficients, also known as loadings, can be expressed in a vector format as follows: $\phi_1 = (\phi_{11}, \phi_{21}, \dots, \phi_{p1})^T$. In order to obtain this linear combination the following optimisations problem is solved, subject to the normalisation constraint: $\sum_{j=1}^p \phi_{j1}^2 = 1$:

$$\max_{\phi_{11}, \phi_{21}, \dots, \phi_{p1}} = \left\{ \frac{1}{n} \sum_{i=1}^n \left(\sum_{j=1}^p \phi_{j1} x_{ij} \right)^2 \right\} \quad (15)$$

This component accounts for the largest variance in the data and it is assumed that the covariates used have been centred to have a mean of zero. The second principal component can then be determined based on it being a linear combination of the covariates that are uncorrelated with Z_1 and it accounting for the second most variation in the data. This process can be continued in this fashion. The Z vectors can now be plotted in a lower dimensional space which can also be understood as a projection the original data down into a subspace spanned by ϕ , this plot is known as a bi-plot. A detailed description of this method can be found in *An Introduction to Statistical Methods* by Trevor Hastie et al.(2013).

This method is used to get an understanding of how each predictor contributes to the overall variability in the data. This is useful to get an understanding of which measures to consider as potential determinants of Sovereign Risk. The principal components themselves are not modelled against the yield measures as this research is primarily interested in establishing and understanding relationships between real macro-economic and political phenomena and the resulting Sovereign Risk.

4.4 | Ordinary Least Squares Regression

An ordinary least squares regression model is fitted, regressing the South African yield spread relative to the respective market yields against the economic and political variables of interest for 10 and 30 year bond yields respectively. The variables of interest can be seen in the *Data Collection and Exploration* section of this study. Effectively a longitudinal analysis is conducted over the period 2012 - 2018 and 2008 - 2016 for the short and long term cases respectively. It is tested whether there is an association between the yield spread differences and the chosen predictors. The OLS regression is formulated as follows;

$$SA\ yield_{k,t} - Market\ yield_{j,k,t} = \beta_0 + \beta X_{k,t} + \epsilon_t$$

where t represents the point in time

$j = \{\text{Developed Markets, Emerging Markets}\}$

and $k = \{10\ \text{year bond, 30 year bond}\}$

(16)

It is assumed that the error terms are normally distributed with a mean of zero and a fixed standard deviation of σ . The independent variables $X_{k,t}$ above consist of the economic and political variables described in Table 8 and Table 9 for the short and long term cases respectively. For the long term case each independent variable is fitted separately. ie multiple univariate distributions are fitted. This is to account for the limited data available in the long term case.

In order to fit the regression models that follow, each series must be checked for stationarity. The stationarity of the respective covariates is checked through the use of visual inspection of their time series plots as well as the augmented Dickey Fuller Unit Root test.

If it is found that a series is not stationary the log difference or normal difference is applied to the data series;

$$x_{stat,t} = \log \left(\frac{x_t}{x_{t-1}} \right)$$

where x represents the covariate series of interest at a time point t

(17)

$$x_{stat,t} = x_t - x_{t-1}$$

where x represents the covariate series of interest at a time point t

(18)

Transformations were applied to the following predictor variables. The initial p-value of the augmented Dickey Fuller Unit Root test is recorded as well as the resulting p-value after the series has been transformed. In Table 11 and 12 it can be seen that the p-values are all significantly small after the transformation of the series.

TABLE 11 Augmented Dickey Fuller test for short term predictors

time series	transfomation	initial p-value	final p-value	p-
Nominal Exchange Rate	log differnced	5.021e-07	2.43e-10	
TED spread	log differenced	0.505	6.609e-09	
Global Premium	-	0.004	0.004	
Market Return	log differenced	0.022	2.2e-16	
EM Sentiment Spread	-	2.263e-08	2.263e-08	
DM Sentiment Spread	-	8.629e-12	8.629e-12	

TABLE 12 Augmented Dickey Fuller test for long term predictors

time series	transfomation	initial p-value	final p-value
Current Account Balance(% of GDP)	Log differenced	0.176	0.075
Foreign Direct Investment	log differenced	0.098	0.0001
Government Effectiveness	log differenced	0.167	0.008
International Debt Issues to GDP	log differenced	0.073	0.012
Political Stability	differenced	0.470	0.014
Regulatory Quality	log differenced	0.999	0.003
Rule of Law	differenced	0.3667	0.001
Tax Revenue (% of GDP)	log differenced	0.380	0.059
Taxes on international trade (% of revenue)	log differenced	0.628	0.027
Voice and Accountability	differenced	0.670	0.017
Total reserves (% of total external debt)	log differenced	0.822	0.010
Total reserves includes gold (% of GDP)	log differenced	0.406	0.007

Note the above transformation essentially converts each series to continuously compounded percentage changes, where the covariate is effectively centred around zeros and divided through by its variance, standardising the variability of the respective measures. Some of the covariates are inherently stationary and thus did not need to be transformed. For

instance the sentiment measures and the Global Premium did not require a transformation. The resulting correlations between the covariates can be seen in Appendix C. It can be seen that there are no significant correlations between the variables.

Once the OLS regression model is fitted, the residual series is assessed to ensure that the residuals are not autocorrelated with a constant variance. These stationarity tests can be seen in Appendix D and E.

4.4.1 | Regression Optimisation Techniques

Each of the techniques explored below are detailed in an *Introduction to Statistical Learning* by Trevor Hastie et al.(2013) [21].

Forward Stage wise, Lasso and Ridge regression

These methods are introduced to improve the predictive ability of the ordinary least squares regression model as well as to mitigate the problems associated with limited sample sizes and to account for correlated covariates. However no significant correlations were found between predictors, see appendix C. Hence the use of these methods is mainly for its dimension reduction capability.

These methods require standardised covariates, where the predictor variables are centred and scaled as follows:

$$x_s = \frac{x - \mu_x}{\sigma_x} \quad (19)$$

where x represents the series of interest

and μ_x and σ_x represent the mean and variance of the series respectively.

Forward stage wise regression is a type of subset selection method. It is a technique that is used to reduce a large number of covariates and hence prevent the overfitting of a model. This method essentially fits a sequence of regression models, with the first model in the sequence being the intercept model. This implies that the coefficients of all the independent covariates is set to zero. The values of these covariates will be updated as one cycles through the algorithm. After the intercept model has been fitted, the variable that is most correlated with the errors of this model is identified and these residuals are then regressed against this variable ($\epsilon_0 = \beta_0 + \beta_1 * x_{\text{highest correlation to } \epsilon_0}$). The resulting coefficient(β_1) of this variable in the regression of the errors is then added to its previous value in the base model (which was zero in the first stage of the algorithm). The base model is then fitted again containing this variable, the errors recorded; ϵ_1 , the most highly correlated variable to the errors; $x_{\text{highest correlation to } \epsilon_1}$ is identified and subsequently the current residuals are regressed against this variable; $\epsilon_1 = \beta_0 + \beta_1 * x_{\text{highest correlation to } \epsilon_1}$ and the resulting coefficient(β_1) added to the base model. This cycle continues in this fashion until none of the inputs are correlated with the residuals or until the model has reached a satisfactory level of fit.

This algorithm is stated more formally as follows; Assuming that the base model has each of its covariates set to zero essentially making it an intercept model. Starting at $k = 0$ follow the steps below;

1. Fit the base model and record errors of this model; ϵ_k
2. Identify covariate that is most highly correlated to residuals of model fitted in step 1.
3. Regress residuals from step 1 against covariate identified in step 2 and record coefficient of covariate.
4. Add coefficient found in step 3 above to the corresponding covariate coefficient value in the base model.
5. Increment k by one and cycle back to step 1, if sufficient fit has still not been achieved.

Conversely shrinkage methods can be applied to the regression model. In this case all of the covariates should be included in the model, as opposed to leaving some variables out as is the case in subset selection. The shrinkage method instead shrinks the size of each coefficient in a continuous manner. This method may be preferred over subset selection as coefficient estimates are unlikely to differ greatly across samples which should result in less sampling variability and lower prediction error. Furthermore Ridge regression is useful in cases where the covariates are highly correlated.

Both Ridge and Lasso regression shrink the regression coefficients by imposing a penalty on their size. This is achieved through minimising the following penalised residual sum of squares (RSS) expression. In the case of Ridge regression $q = 2$ while for Lasso regression $q = 1$. Furthermore the case in-between Lasso and Ridge is represented by Elastic Net regression with $q = 0.5$.

$$\hat{\beta}_{ridge} = \underset{\beta}{\operatorname{argmin}} \left\{ \sum_{t=1}^N (y_{i,t} - \beta_0 - \sum_{j=1}^p x_{t,j} \beta_j)^2 \right\}$$

(20)

such that $\sum_{j=1}^p |\beta_j|^q \leq s$

$s \geq 0, q = 1, 2, p = \text{number of predictor variables}, i = \text{country of market of interest}.$

Where s refers to a budget for how large $\sum_{j=1}^p |\beta_j|^q$ can be. If s is large then the budget is not very restrictive, allowing the coefficient estimates to remain large. If s is extremely large, one will find that the coefficients are that of the OLS regression. If s is small the coefficients must shrink in order to satisfy this constraint. Lastly N refers to the number of observations within the sample to which the model is being fitted.

In Lagrangian form the above can be expressed as:

$$\hat{\beta}_{ridge} = \underset{\beta}{\operatorname{argmin}} \left\{ \sum_{t=1}^N (y_t - \beta_0 - \sum_{j=1}^p x_{t,j} \beta_j)^2 + \lambda \sum_{j=1}^p |\beta_j|^q \right\}$$

(21)

$\lambda \geq 0, q = 1, 2$

$p = \text{number of predictor variables}, N = \text{number of timepoints}.$

In the case of Ridge regression the above can be expressed in matrix notation as follows:

$$\hat{\beta}_{ridge} = \begin{cases} (X^T X + \lambda I)^{-1} X^T \underline{y} & \text{for all estimates excluding the intercept} \\ \bar{y} & \text{for the intercept estimate} \end{cases} \quad (22)$$

If the covariates are highly correlated, the matrix X of covariates is singular (non-invertible), since the eigenvectors are multiples of each other resulting in a matrix that is not of full rank. Ridge regression however accounts for this problem by adding a λ value to the diagonal of the symmetric $X^T X$ matrix, making it of full rank and hence invertible. However as a result of the added λ value, Ridge regression coefficients become biased, unless $\lambda = 0$. It can be seen that the bias is in fact monotonically increasing in λ . Furthermore it is noted that the total variance of the coefficients is conversely monotonically decreasing in λ :

$$\begin{aligned} Bias(\hat{\beta}_{ridge}) &= -\lambda(X^T X + \lambda I)^{-1} \beta \\ Var(\hat{\beta}_{ridge}) &= \sigma^2 W X^T X W \\ \text{where } W &= (X^T X + \lambda I)^{-1} \end{aligned} \quad (23)$$

It is clear that as λ tends to zero the Ridge regression coefficient estimates will tend to that of the ordinary least squares estimate. Furthermore the optimal value of lambda can be determined through the use of cross validation, where one can pick the value of λ that minimises the expected prediction error.

4.4.2 | Short-Term Analysis

For the short term analysis (10 year bond yield) variables recorded at a monthly or daily frequency are considered. These include the developed Sentiment Scores, the Global Premium, Market Index Returns, the Exchange Rate and the TED spread. Variables recorded on a daily basis have been aggregated to a monthly basis. Matching the predictor and response variables, data for the period 2012 - 2018 is obtained. The regression formulation is as follows, where the Market Yield is that of either the Emerging or Developed Markets. Note each spread measure is relative to US yield.

$$\begin{aligned} SA\ yield_t - Market\ yield_{j,t} = & \\ & \beta_0 + Developed\ Markets\ Sentiment\ Score_t * \beta_1 + Emerging\ Markets\ Sentiment\ Score_t * \beta_2 \\ & + Global\ Premium_t * \beta_3 + Market\ Return_t * \beta_4 + Exchange\ Rate_t * \beta_5 + TED\ Spread_t * \beta_6 + \epsilon_t \end{aligned} \quad (24)$$

where t represents the month & year

$j = \{\text{Developed Markets, Emerging Markets}\}$

ϵ_t the residuals of the model at time point t .

4.4.3 | Long-Term Analysis

The long term analysis (30 year bond yield) is conducted on a yearly frequency. After matching the available data of the respective predictors against the yield spread measure, the resulting data ranges over the period 2008 - 2016. The formulation of the respective univariate models is as follows, with the Economic Indicators and Political Indicators being those outlined in the *Data Collection & Exploration* section.

$$\begin{aligned}
 SA\ yield_t - Market\ yield_{j,t} = \\
 \beta_0 + Economic/Political\ Indicator_t * \beta_1 + \epsilon_t
 \end{aligned}
 \tag{25}$$

where t represents the year

$j = \{\text{Developed Markets, Emerging Markets}\}$

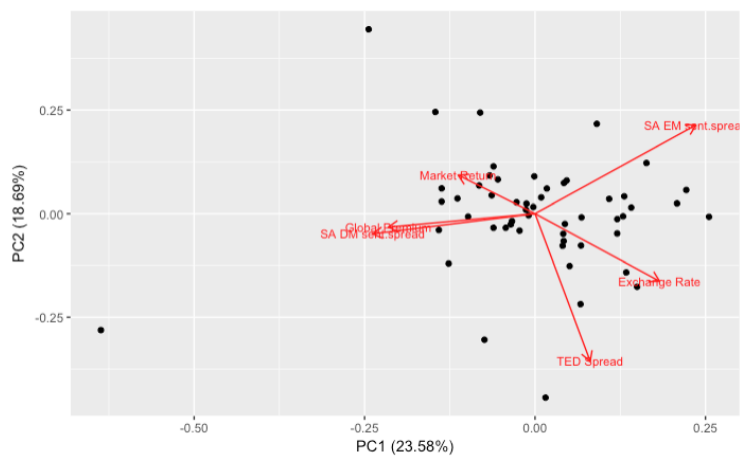
ϵ_t the residuals of the model at time point t .

5 | MODELLING, RESULTS & ANALYSIS

5.1 | PCA

The PCA conducted below is based on scaled and centred covariates for the respective analyses. Scaling the data allows one to compare the contribution to variability of the respective variables on the same scale. Conducting PCA on the monthly and yearly covariates respectively the following plots of the respective principal components are obtained;

FIGURE 18 Short term covariate PCA plot for the period 2012 - 2018

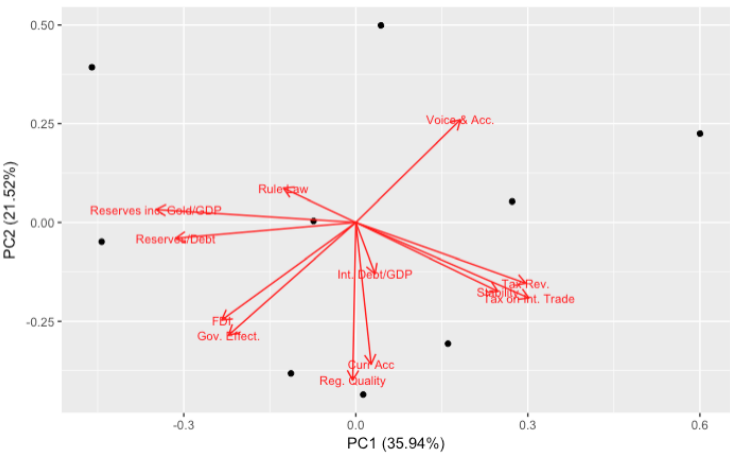


For the short term case it can be seen that the Sentiment Spread measures contribute mostly to the first component, while the second Principal Component is dominated by the TED Spread and the Emerging Markets Sentiment Spread variables. This interpretation can be further substantiated through the Principal Component loading vectors in Table 13, where the larger coefficients for the aforementioned variables can be seen.

TABLE 13 Loading Vectors for the respective Principal Components found for the Short Term Analysis case.

Covariate	PC1	PC2
Exchange Rate	0.397	-0.355
Ted Spread	-0.176	-0.775
SA DM sent. spread	-0.518	-0.106
SA EM sent. spread	0.513	0.465
Global Premium	-0.468	-0.071
Market Return	-0.245	0.202

FIGURE 19 Long Term covariate PCA plot for the period 2008 - 2016



For the long term case it can be see that the first Principal Component is dominated by the respective reserve measures: Reserves including Gold (% of GDP) and Reserves (% of external debt). The second Principal Component is dominated by the Current Account (% of GDP) and Regulatory Quality variables. These insights are substantiated by the loading vectors displayed in Table 14 below.

TABLE 14 Loading Vectors for the respective Principal Components found for the Long Term Analysis case.

Covariate	PC1	PC2
Curr. Acc.	0.034	-0.458
FDI	-0.298	-0.315
Gov. Effect.	-0.283	-0.363
Int. Debt/GDP	0.043	-0.166
Stability	0.316	-0.222
Reg. Quality	-0.007	-0.509
Rule Law	-0.161	0.111
Tax Rev.	0.378	-0.196
Tax on Int. Trade	0.385	-0.244
Reserves/Debt	-0.400	-0.051
Reserves inc. Gold/GDP	-0.444	0.042
Voice & Acc.	0.232	0.331

5.2 | Regression

5.2.1 | Short Term Analysis

For the short-term analysis case (10-year bonds), the respective models are fitted on a training set, accounting for 90% of the collected data. The model fits are then compared on a validation set. The robustness of the model fit is then assessed through the validation set error.

OLS Regression

Given the considerable number of data points (53 observations ranging from 2012-2018), one is able to fit an OLS regression using the entire set of available predictors. In this case there are a total of 6 predictors that are tested for a potential association to the response.

Fitting a full OLS regression model to the South African yield spreads to the respective markets Sovereign Yield the following regression summary statistics are calculated in Table 15. The error plots for this model can be found in *Appendix D*;

From equation 26 below it can be seen that the South African yield spread to the Emerging Markets yield has been regressed against the following predictors; Global Premium, Market Return, Exchange Rate, Ted Spread, Developed Market Sentiment Spread and Emerging Market Sentiment Spread score. The residuals of this model are recorded, ranging from -1.433 to 0.872. This range is slightly negatively skewed. Looking at the coefficients of the model, under the estimate column in Table 15, it can be seen that only the Global Premium and Market Return are positively associated with the Sovereign Risk measure. All the other covariates; The Exchange Rate, Ted spread, Developed Market Sentiment and Emerging Market Sentiment are negatively associated with the Sovereign risk measure. Furthermore looking at the standard error of the coefficients one is able to assess the hypothetical variance of the given coefficient if the model were fitted to the data multiple times. For each of the covariates the standard error is larger than the covariate itself, effectively indicating that the given associations found could easily change in sign and are not significantly different from zero. The t-test is then applied to each of the predictors to assess whether the association between the independent and dependent variable is significant. It can be seen that the p-values associated with the respective t-statistics, under the $Pr(> |t|)$ column in Table 15, are all large indicating that the chosen covariates are not significantly associated with the Emerging Market yield spread.

$$\begin{aligned}
 SA \text{ Market Yield Spread} - Emerging \text{ Markets Yield Spread} = & \\
 - 1.002e - 01 + 4.328e - 02 * Global \text{ Premium} + 5.979e - 01 * Market \text{ Return} & \\
 - 1.002e - 01 * Exchange \text{ Rate} - 2.387e - 02 * Ted \text{ Spread} & \\
 - 2.133e - 05 * SA \text{ DM Sentiment Spread} - 1.489e - 03 * SA \text{ EM Sentiment Spread} &
 \end{aligned}
 \tag{26}$$

TABLE 15 Regression Summary Statistics for Emerging Markets yield spread.

Residuals:				
Min	1Q	Median	3Q	Max
-1.433	-0.260	0.175	0.443	0.872
Coefficients:	Estimate	Std.Error	t value	Pr(> t)
(Intercept)	-1.002e-01	5.362e-01	-0.187	0.853
Exchange Rate	-1.002e-01	5.362e-01	-0.187	0.853
Ted Spread	-2.387e-02	1.055e-01	-0.226	0.822
SA DM Sentiment Spread	-2.133e-05	5.213e-04	-0.041	0.968
SA EM Sentiment Spread	-1.489e-03	2.471e-03	-0.602	0.550
Global Premium	4.328e-02	2.538e-01	0.171	0.865
Market Return	5.979e-01	2.881e+00	0.208	0.837

Significance codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1				
Residual Standard Error: 0.6166 on 40 degrees of freedom				
Multiple R-Squared: 0.01738, adjusted R-squared: -0.13				
F-statistics: 0.1179 on 6 and 40 DF, p-value:0.9937				

It is evident that in the short run the Sovereign Yield spread between South Africa and the Emerging Markets cannot be accurately determined using the above predictors. There is not enough evidence against the null hypothesis, with a p-value of 0.99. It can also be seen that the overall fit of the model is poor with an adjusted R-squared, which indicates how well the model is fitting the data, of -0.13. This negative value indicates that after the adjustment for the number of variables included in the model, the R-squared metric actually becomes negative.

Fitting this model on the validation set, a mean square error of 0.76 is obtained.

An OLS regression model is then also fitted to the South African Sovereign yield and the Developed Markets yield spread. In Table 16 it can be seen that the residuals of the model include zero within the range of values and are negatively skewed. A significant relationship is evident between the yield spread and the Global Premium, where one unit increase in the Global Premium (difference between S&P 1200 return and US yield) is associated with a decrease in the yield spread of 0.93 percent, as indicated by equation 27. This relationship indicates that an increase in the global risk environment is associated with a decrease in the Sovereign Risk difference between South Africa and the Developed Markets. Essentially implying that in an uncertain economic environment the difference in Sovereign Risk in South Africa and the Developed Markets becomes slightly less.

$$SA\ Market\ Yield\ Spread - Developed\ Markets\ Yield\ Spread =$$
$$- 0.895 - 0.934 * Global\ Premium + 3.198 * Exchange\ Rate - 0.054 * Ted\ Spread$$
$$- 0.0008 * SA\ DM\ Sentiment\ Spread + 0.0008 * SA\ EM\ Sentiment\ Spread - 0.081 * Market\ Return$$

(27)

TABLE 16 Regression Summary Statistics for Developed Markets yield spread.

Residuals:				
Min	1Q	Median	3Q	Max
-1.456	-0.363	-0.009	0.329	1.167
Coefficients:	Estimate	Std.Error	t value	Pr(> t)
(Intercept)	-0.895	0.534	-1.676	0.102
Exchange Rate	3.198	3.161	1.012	0.318
Ted Spread	-0.054	0.105	-0.510	0.613
SA DM Sentiment Spread	-0.0008	0.0005	-1.553	0.128
SA EM Sentiment Spread	0.0008	0.002	0.328	0.744
Global Premium	-0.934	0.253	-3.698	0.0006 ***
Market Return	-0.081	2.868	-0.028	0.978

Significance codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1				
Residual Standard Error: 0.614 on 40 degrees of freedom				
Multiple R-Squared: 0.3735, adjusted R-squared: 0.2795				
F-statistics: 3.974 on 6 and 40 DF, p-value: 0.00328				

Overall the fit of the data is still low with only 28% of the variation in the data being accounted for (adjusted R-squared of 0.279). However the overall significance of the regression is 0.003 (p-value), indicating that there is indeed a significant relationship between the SA yield Spread to the Developed Markets Yield and the chosen predictor variables.

Fitting this model on the validation set, produces a mean square error of 0.13. This is considerably lower compared to the Emerging Markets case, which has a MSE of 0.76.

Forward Stage-wise Regression

Applying the forward-stagewise algorithm to the short-term analysis case, the following stages were obtained with RSS at each stage as follows;

TABLE 17 Emerging Markets Case: RSS for each stage of the Forward Stage-wise algorithm.

Stage	covariate added	RSS	% change in RSS
0	Intercept	15.478	-
1	SA EM Sentiment Spread	15.378	-0.65
2	Exchange Rate	15.312	-0.43
3	Global Premium	15.303	-0.06
4	Market Return	15.296	-0.05
5	TED Spread	15.211	-0.56
6	SA MM Sentiment Spread	15.209	-0.01

For the Emerging Markets comparison of Sovereign Risk it can be seen that the most significant reduction in RSS occurs in the first 2 stages of the algorithm, see percentage changes in RSS in Table 17 above, when the Emerging Market Sentiment Spread and Exchange Rate is inserted into the model. An OLS regression of the South African EM Yield Spread regressed against these predictors produces an adjusted R-Squared of -0.031 (previously -0.13) with a p-value of 0.731. This is an improvement over the full model which had a p-value of 0.994. However the overall fit is still poor.

Using this model to predict the validation set response, a MSE of 0.76 is achieved.

TABLE 18 Developed Markets Case: RSS for each stage of the Forward Stage-wise algorithm.

Stage	covariate added	RSS	% change in RSS
0	Intercept	24.07	-
1	Global Premium	17.96	-25.38
2	SA MM Sentiment Spread	16.7	-6.71
3	Exchange Rate	15.69	-6.35
4	SA EM Sentiment Spread	15.37	-2.03
5	Ted Spread	15.081	-1.89
6	Market Return	15.079	-0.01

For the Developed Markets case, it can be seen that the most significant improvement is in the first stage of the algorithm, see Table 18. The model at this stage contains the Global Premium. Fitting an OLS regression to include only this variable an adjusted R-squared value of 0.294 and a p-value of 4.993e-05 is achieved. A significant relationship between the yield spread and the Global Premium is found, with a unit increase in the Global Premium associated with 1.05 decrease in the yield spread. Fitting this model to the validation set a MSE of 0.127 is achieved, this is an improvement in fit over the full OLS regression model (MSE of 0.13).

Lasso Regression

Applying Lasso regression one first needs to establish the value of λ , that will effectively minimise the expected prediction error. The resulting value of lambda is then fed into equation 21 which is minimised to obtain the optimal fit for the regression model. A range of λ values are tested. For each λ value that is tested within this range an expected prediction error value is obtained. The λ corresponding to the smallest error is optimal. The prediction-error value is determined through the use of 5-fold cross validation where the data is effectively separated into 5-folds and each fold's yield spread is estimated based on fitting the model to the other 4-folds and then using that model to estimate the yield spread of the current target fold of data.

In figure 20 and 21 below one can see the expected error values associated with each value of lambda for the Emerging and Developed Market models. To avoid overfitting the model the largest lambda value within 1 standard deviation of λ_{min} is used as the optimal lambda. The number of predictors included in the model can be read off the top horizontal axis of the graphs below. Predictors are essentially excluded from the model when their coefficients are shrunk all the way to zero.

FIGURE 20 Expected Lasso Prediction Error for tested range of lambda values for Emerging Markets yield spread.

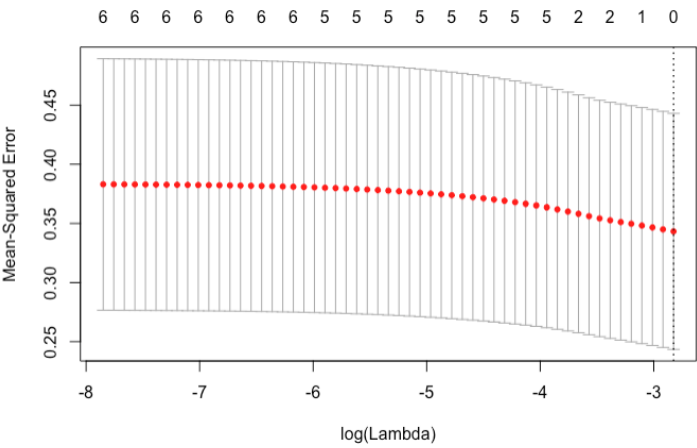
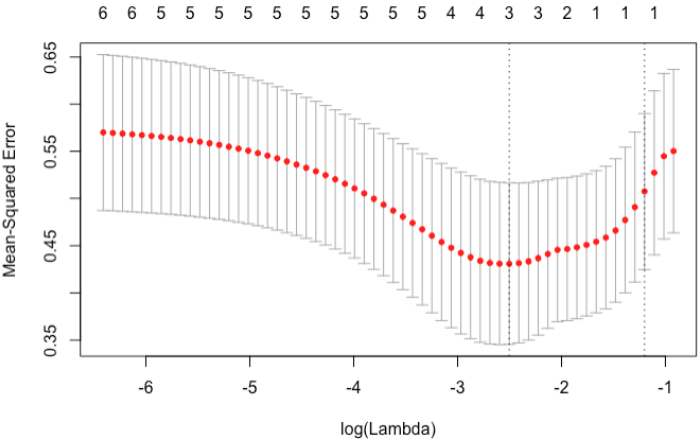


FIGURE 21 Expected Lasso Prediction Error for tested range of lambda values for Developed Markets yield spread.



From the above Cross Validation (CV) graphs the optimal values of lambda can be calculated as 0.059 and 0.30 for the Emerging and Developed Markets respectively. These being the values of lambda within one standard deviation at which the MSE is minimised. The model is subsequently fitted with these values of lambda and when tested on the validation set the following MSE are recorded; 0.725 and 0.194 for the Emerging and Developed Markets respectively. Furthermore it can be read off the top horizontal axis that the optimal number of variables to include in the model is 1 and 3 respectively.

Notice from the above graphs that the error is in fact decreasing with fewer variables included in the model. This phenomena would not be possible if the MSE was calculated in sample. In this case it is calculated out of sample on the respective target folds using CV. The MSE formula associated with this method is as follows;

$$M\hat{S}E = b\hat{a}s^2 + v\hat{a}r \tag{28}$$

As the number of variables included in a model increase, the bias decreases monotonically. However $v\hat{a}r$ is monotonically increasing with the number of variables included in the model. The above error plots indicate that as variables are excluded or shrunk in the model the reduction in variability outweighs the increase in bias of the model, resulting in a lower MSE as seen above.

Elastic Net Regression

For Elastic Net regression the optimal value of lambda also needs to be determined.

FIGURE 22 Expected Elastic Net Prediction Error for tested range of lambda values for Emerging Markets yield spread.

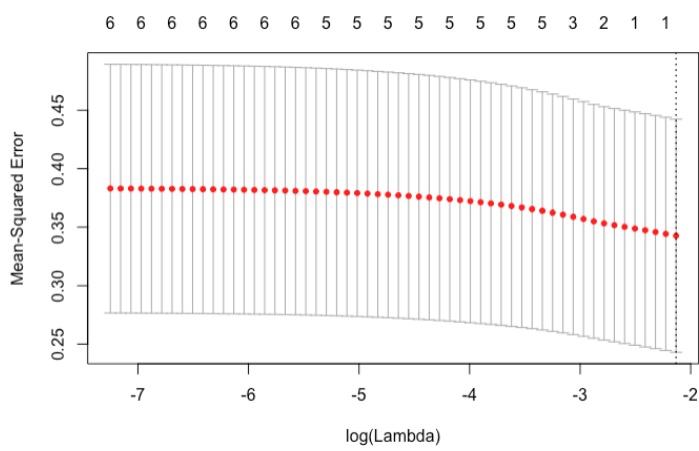
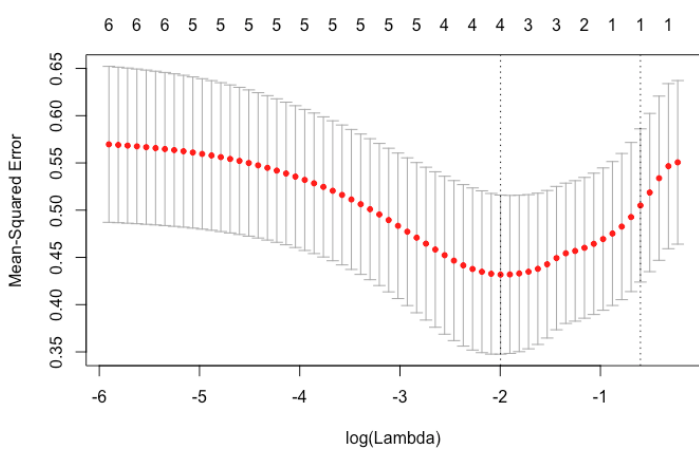


FIGURE 23 Expected Elastic Net Prediction Error for tested range of lambda values for Developed Markets yield spread.



From these CV plots the optimal values of lambda are calculated as 0.119 and 0.548 for the Emerging and Developed Markets. Furthermore the optimal number of variables to include in the model is 1 and 4 respectively. Fitting the models with this value of lambda and subsequently testing the fit on the validation set a MSE of 0.725 and 0.200 is achieved.

Ridge Regression

FIGURE 24 Expected Ridge Prediction Error for tested range of lambda values for Emerging Markets yield spread.

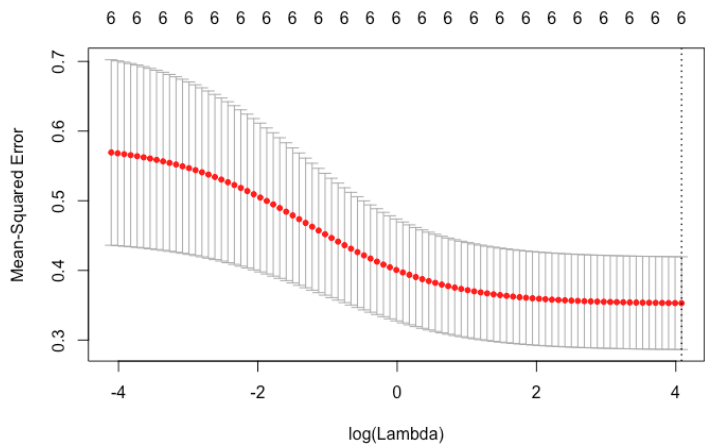
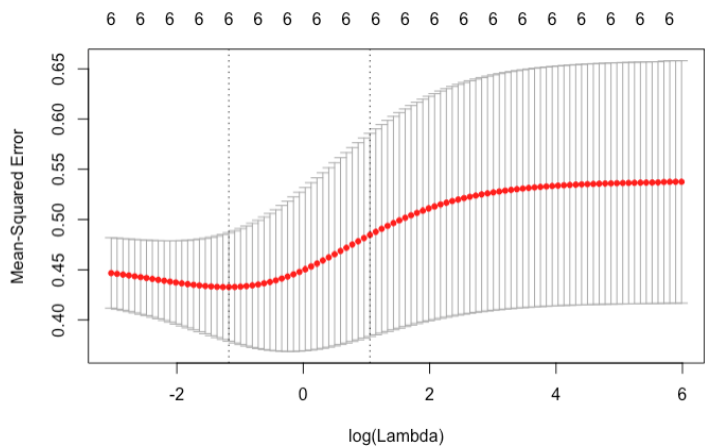


FIGURE 25 Expected Ridge Prediction Error of tested range of lambda values for Developed Markets yield spread.



For the Ridge regression case an optimal value of lambda is obtained at 59.273 and 2.872 for the Emerging and Developed Markets. The resulting models produce a MSE of 0.726 and 0.222 on the validation set. For both Ridge regression models all of the predictors remain in the model.

Regression Optimiser Comparison

Having established the relative fitting capabilities of the above regression optimisation techniques, one is now able to

select the model that best fits the data. A summary of the respective model performances on predicting the validation set response is given below. It can be seen that overall the Developed Markets model has produced smaller MSEs. It can also be seen that the OLS regression model produced the best results, with the smallest MSE, for the Developed Markets case. For the Emerging Markets case the Lasso and Elastic Net regression method produced the lowest MSEs.

TABLE 19 Regression results

Short Term Regression Results: MSE		
Model	Emerging Markets	Developed Markets
OLS	0.76	0.13
Forward Stagewise	0.76	0.13
Lasso	0.73	0.20
Elastic Net	0.73	0.20
Ridge	0.73	0.22

5.2.2 | Long Term Analysis

The analysis below is conducted on the entire set of the data. Due to the limited amount of available data, only 9 data points , univariate OLS Regression models are fitted to avoid overfitting the data. The corresponding p-values are recorded in order to compare the relationships found between the yield spread and the respective determinants. The residuals of the respective models are tested for stationarity, results for which can be found in appendix E.

OLS Regression

TABLE 20 Regression results

Long Term Regression Results: p-value		
Fitted Predictor	Emerging Markets	Developed Markets
Current account balance (% of GDP)	0.664	0.646
Foreign direct investment, net inflows (% of GDP)	0.141	0.130
Government Effectiveness	0.280	0.147
International debt issues to GDP (%)'	0.690	0.927
Political Stability	0.172	0.448
Regulatory Quality	0.990	0.718
Rule of Law	0.996	0.829
Tax revenue (% of GDP)	0.149	0.706
Taxes on international trade (% of revenue)	0.034	0.102
Total reserves (% of total external debt)	0.008	0.068
Total reserves includes gold (% of GDP)	0.002	0.446
Voice & Accountability	0.082	0.597

From the above results it is clear that for the Emerging Market yield spread, the most significant relationship exists between the yield spread and the following predictors: Total reserves includes gold (% of GDP), Total reserves (% of total external debt) and Taxes on international trade (% of revenue). For these variables the following relationships are established:

- A unit increase in the Total reserves includes gold (% of GDP), is associated with a 9.35 increase in the yield spread.
- A unit increase in the Total reserves (% of total external debt) , is associated with a 6.527 increase in the yield spread.
- A unit increase in Taxes on international trade (% of revenue) , is associated with a 3.323 decrease in the yield spread.

In the developed market case is can be seen that the yield spread is most significantly related to Total reserves (% of total external debt), Taxes on international trade (% of revenue) and Foreign direct investment, net inflows (% of GDP). These relationships are quantified as follows:

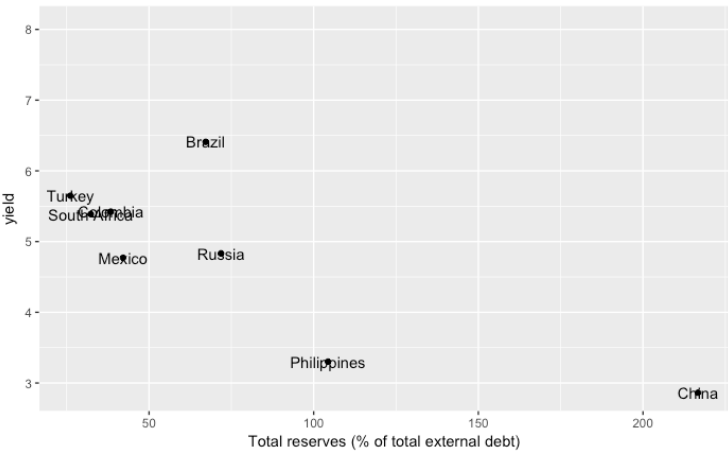
- A unit increase in the Total reserves (% of total external debt), is associated with a 7.247 increase in the yield spread.
- A unit increase in the Taxes on international trade (% of revenue), is associated with a 3.905 decrease in the yield spread.
- A unit increase in the Foreign direct investment, net inflows (% of GDP), is associated with a 0.931 increase in the yield spread.

5.3 | Results Visualisation

Given the significant results of the South African Sovereign Risk measure to the Emerging Markets in the long term, scatter plots are utilised to further understand which countries within the Emerging Markets are responsible for the difference in Sovereign Risk. The relationship between the yield and the respective predictors is taken as of 2016. The significant predictors were found to be the respective reserves measures: Total reserves includes gold (% of GDP), Total reserves (% of total external debt) and Taxes on international trade (% of revenue).

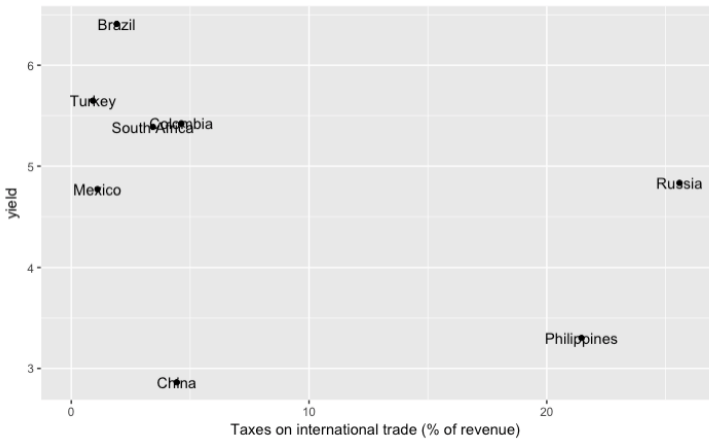
With respect to reserves South Africa is very closely clustered to Turkey, Columbia and Mexico (see figure 26), while particularly far away from China which has a much smaller yield and larger reserve pool.

FIGURE 26 Total reserves (% of total external debt) and the corresponding 30-year yield for the respective Emerging Markets countries.



Looking at taxes on international trade South Africa is again clustered with Turkey, Mexico and Columbia. Russia and the Philippines are seen to be outliers with larger taxes as a % of revenue and smaller yields in comparison to South Africa.

FIGURE 27 Taxes on international trade and the corresponding 30-year yield for the respective Emerging Markets countries.



6 | DISCUSSION OF RESULTS

It is evident that the ability to ascertain drivers of the difference in risk across South Africa and Emerging and Developed Markets respectively is different across the short and long run. In the short run there was no significant relationship found between the Emerging Market yield spread and the chosen predictors. However movements in the Developed Market yield spread were found to be associated to the Global Premium, where an increase in the global premium is associated with a decrease in the yield spread. In the short run for both the Emerging and Developed Market models only a limited amount of variability in the data is accounted for with the OLS models explaining only -0.13 and 0.279 of the variation in the data respectively.

In the long run South Africa's Sovereign Risk spread to the Emerging Markets is associated with the movements in the following predictors: Total reserves includes gold (% of GDP), Total reserves (% of total external debt) and Taxes on international trade (% of revenue). An increase in the reserve predictors, and hence in the stability of the South African economy, is associated with an increase in the yield spread between South Africa and the Emerging Markets. While an increase in tax on international trade is associated with a decreased yield spread. For the Developed Market comparison it is found that the level of reserves is strongly associated to the difference in risk across the markets, with an increase in South Africa's level of reserves associated with an increase in the yield spread. Furthermore Taxes on international trade (% of revenue) is also found to be significant, with an inverse relationship being evident. This is a similar relationship that was found in the Emerging Markets case. Furthermore FDI is also found to be significant in understanding the difference in risk environments across South Africa and Developed Markets, where an increase in FDI in South Africa is associated with an increase in the yield spread.

In terms of model improvement through the use of the optimisation techniques, only limited improvements in fit were able to be achieved. However these techniques did provide a more intensive interrogation of which predictors could

be relevant. In the Forward-Stage wise algorithm the importance of the Sentiment Spread and the Exchange Rate is highlighted for both the Emerging and Developed Market cases, with these being the first two variables chosen to include in the respective models.

The developed Sentiment measures performance did not prove to have a significant association with the Sovereign Risk spread measure. It was hoped that this measure would accurately measure the sentiment of the market, which is in turn linked to the risk associated with a given market. In future work the measure may be improved in terms of how it is calculated. Furthermore smoothing techniques could be implemented to reduce the variability in the measure. A smoother measure may be able to track the risk in the economy more accurately.

7 | GLOSSARY

- CV: Cross Validation
- EM: Emerging Market
- GFC: Global Financial Crisis
- DM: Developed Market

Appendices

A | YIELD DATA MANIPULATION

The sequence of yield data manipulations are described below as well as the corresponding *YieldAnalysis* functions that were used to perform these actions.

1. The data is first collected from the Bloomberg terminal. Each bond forms its' own csv file and is uploaded into R. From here the individual series are cleaned (*data_clean*) to rid the series of unnecessary spaces as well as to format the dates. The cleaned series are then merged together (*bond_merger*) to form a table of yield series for each sampled country.
2. The number of days to maturity can now be calculated for each bond (*DTM* & *DTM_merger*). Where the merger function is able to calculate time-to-maturities for a list of bonds given the relevant date series of the bond yields. The days are then converted to months, based on there being approximately 30 days in a month, as well as to years, based on there being approximately 365 days in a year.
3. The matching process is now conducted using the country yield tables and time-to-maturity tables created previously.
 - For each bond, information on its yield and time-to-maturity is merged. This results in a table of yields corresponding to all the unique time-to-maturity and date combinations(*indiv_bonds*).
 - These individual bond tables are then merged according to country. An average yield value is calculated in the cases where there is more than one bond with the same time-to-maturity and date belonging to the same country. This results in one series of yield values belonging to each country (*Country_AvgYieldTTM*).
 - The yield series of each country is finally merged to form the final data frame.

B | YIELD SPREAD DETERMINANTS

TABLE 21 Yield Spread Determinants, their descriptions and data sources

Determinant	Description	Source
Economic Variables		
Global Equity Premium	Index that measures the global equity market performance	Bloomberg terminal
Local Equity Premium	Country specific market returns	Bloomberg terminal
Nominal Exchange Rate	Quoted in rands per dollar	South African Reserve Bank
TED Spread	This is an interest spread measure between the LIBOR (interest rate on interbank loans) interest rate and U.S. treasury interest rate. It is used to gage global liquidity conditions.	Bloomberg terminal
International Debt Issues to GDP(%)	A measure of the international debt level relative to the country's GDP	World Bank
Current Account balance (% of GDP)	A measure of the current account balance relative to the country's GDP	World Bank
Foreign Direct Investment (FDI), net inflows (% of GDP)	A measure of the FDI relative to the country's GDP	World Bank
Tax Revenue (% of GDP)	A measure of tax revenue collected relative to the country's GDP	World Bank
Taxes on international trade (% of revenue)	A measure of international tax revenue collected relative to the country's total revenue	World Bank
Total Reserves includes gold (% of GDP)	A measure of reserves relative to the country's GDP	World Bank
Total Reserves (% of total external debt)	A measure of total reserves relative to the country's total external debt	World Bank

Political Stability		
Political Stability Index	Index that measures the probability that the government will be destabilised or overthrown. This includes domestic violence and terrorism.	Worldwide Governance Indicators, World Bank
Quality of Governance		
Rule of law	Index that measures the extent to which agents have confidence in and abide by the rules of society. Particularly the quality of contract enforcement, property rights, the police, the courts and the likelihood of crime and violence. The higher this value the more effective the legal system.	Worldwide Governance Indicators, World Bank
Regulatory quality	Measures the governments ability to formulate and implement policies and regulations. Particularly those that promote private sector development. The higher the value the better the regulatory quality.	Worldwide Governance Indicators, World Bank
Government effectiveness	Measures the quality of public and civil services as well as the degree of independence from political pressures. Furthermore also considers the quality of policy formulation and implementation, the credibility of government commitment to those policies. The higher the value the more effective the policies.	Worldwide Governance Indicators, World Bank
Freedom from corruption	Measures the extent to which public power is perceived to be used for personal gain. This includes small and large scale corruption as well as capture of the state by elites. The higher the value the lower the perceived corruption level.	Worldwide Governance Indicators, World Bank
Voice and accountability	Measures the extent to which citizens can participate in selecting their government. Also measures the freedom of expression, association and media. The higher the value, the more democratic rights enjoyed by citizens.	Worldwide Governance Indicators, World Bank

C | PREDICTOR CORRELATIONS

FIGURE 28 Correlation between short term analysis predictors.

	Avg_TED	SA_MM_sent.spread	SA_EM_sent.spread	Global_premia	Exchange_Rate	return
Avg_TED	1.00000000	-0.087011790	-0.11207039	-0.050805562	0.139486808	-0.025497174
SA_MM_sent.spread	-0.08701179	1.000000000	-0.19536005	0.156278838	-0.002973056	0.101663186
SA_EM_sent.spread	-0.11207039	-0.195360046	1.00000000	-0.134528157	0.158938075	-0.021667483
Global_premia	-0.05080556	0.156278838	-0.13452816	1.000000000	-0.083145707	0.008705755
Exchange_Rate	0.13948681	-0.002973056	0.15893807	-0.083145707	1.000000000	-0.074069336
return	-0.02549717	0.101663186	-0.02166748	0.008705755	-0.074069336	1.000000000

FIGURE 29 Correlation between long term analysis predictors.

	Curr Acc Bal	FBI Gov. Effect.	Int Debt-GDP	Political Stab	Reg Qual	Rule Low	Tax Rev % GDP	Tax on international trade	Total Reserves % external debt	Total Reserve % GDP	Voice
Curr Acc Bal	1.00000000	0.34303093	0.111090928	0.204767902	0.00288422	0.53461132	-0.4202938	0.4411278	0.15117088	-0.06035843	0.008009863
FBI	0.343030926	1.00000000	0.829463551	0.016375317	-0.05155229	0.18218255	0.4355468	-0.2182010	-0.38990148	0.43849149	0.392674184
Gov. Effect.	0.111090928	0.82946355	1.000000000	-0.003990053	0.06616216	-0.46699857	0.4822384	-0.2051867	-0.17089881	0.57534631	0.315987868
Int Debt-GDP	0.204767902	0.01637532	-0.003990053	1.000000000	0.08166365	0.02168523	-0.2280207	0.1989578	0.15802581	-0.06665451	0.119674644
Political Stab	0.002884220	-0.05155229	0.066162164	0.081663655	1.00000000	0.11514381	0.1397806	0.5460117	0.79355998	-0.47233422	-0.79179335
Reg Qual	0.534611318	0.18218255	-0.466998570	0.021685230	0.11514381	1.00000000	-0.2081667	0.1847653	0.58714495	0.17789316	-0.851515178
Rule Low	-0.420293790	0.43554677	0.482238413	-0.228020716	0.13978061	-0.20816672	1.0000000	-0.4091728	-0.14130398	0.22446142	0.056492580
Tax Rev % GDP	0.441127843	-0.21820097	-0.205186696	0.198957760	0.54601166	0.18476533	-0.4091728	1.0000000	0.53853725	-0.49176664	-0.756736715
Tax on international trade	0.151170878	-0.38990148	-0.176898809	0.158025809	0.79355998	0.58714495	-0.1413040	0.5385372	1.00000000	-0.65588220	-0.787864750
Total Reserves % external debt	-0.060358426	0.43849149	0.575946315	-0.066654508	-0.47233422	0.17789316	0.2244614	-0.4917666	0.65588220	1.00000000	0.750798362
Total Reserve % GDP	0.008009863	0.39267410	0.315987868	0.119674644	-0.79179333	-0.85151518	0.0564925	-0.7567367	-0.78786475	0.75079836	1.000000000
Voice	-0.311262548	-0.41106549	-0.399825549	-0.131003246	0.07160659	-0.38654922	0.1504234	0.4773651	-0.04754544	-0.29937711	-0.520785302

D | ERROR PLOTS FOR SHORT TERM ANALYSIS

FIGURE 30 Short Term Residual Plot for OLS regression on entire set of predictors: Emerging Markets Case

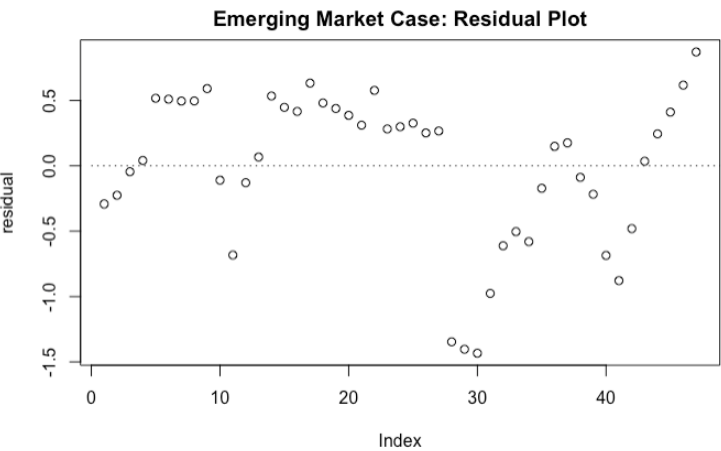
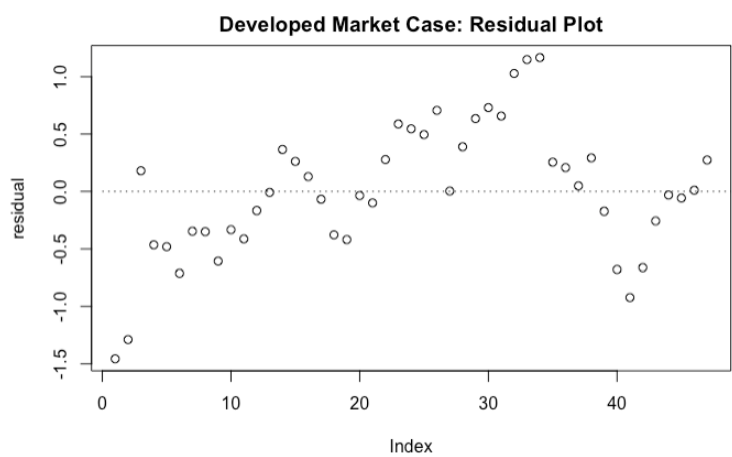


FIGURE 31 Long Term Residual Plot for OLS regression on entire set of predictors: Developed Market Case



E | STATIONARITY OF ERRORS IN LONG TERM ANALYSIS

TABLE 22 Regression results

testing Long Term Regression Residuals for stationarity using the Augmented Dickey Fuller test	
Fitted predictor	p-value
Current account balance (% of GDP)	0.233
Foreign direct investment, net inflows (% of GDP)	0.263
Government Effectiveness	0.187
International debt issues to GDP (%)	0.237
Political Stability	0.140
Regulatory Quality	0.252
Rule of Law	0.244
Tax revenue (% of GDP)	0.208
Taxes on international trade (% of revenue)	0.056
Total reserves (% of total external debt)	0.109
Total reserves includes gold (% of GDP)	0.147
Voice & Accountability	0.200

REFERENCES

- [1] Investments FS. What is sovereign risk and how to measure it. First State Investments; 2010.
- [2] Leadenhall. Is Protectionism on the rise in Australia? Leadenhall; 2017.
- [3] Izadi S, Hassan K M. Impact of international and local conditions on sovereign bond spreads: International evidence. *Borsa Istanbul Review* 2018;18(1):41–51.
- [4] Tan R, The U.S.- China trade war has begun. Here's how things got to this point.; 2018. https://www.washingtonpost.com/news/worldviews/wp/2018/07/05/a-timeline-of-how-the-u-s-china-trade-war-led-us-to-this-code-red-situation/?noredirect=on&utm_term=.5e8dafad5431. Washington Post.
- [5] Hunt A, Wheeler B, Brexit: All you need to know about the UK leaving the EU.; 2018. <https://www.bbc.com/news/uk-politics-32810887>. BBC.
- [6] Dimitriou D, Kenourgios D, Simos T. Global financial crisis and emerging stock market contagion: A multivariate FIAPARCH-DCC approach 2013;.
- [7] ; 2018. <https://bizfluent.com/info-10002682-difference-between-developing-countries-emerging-countries.html>.
- [8] Economist T. Whats in a name? Definition. The Economist; 2017.
- [9] Fund IM. China's Economic Outlook in 6 Charts. International Monetary Fund; 2018.
- [10] Ababa A, China's Industrialization: Overview; 2014.
- [11] Stieglitz S, Dang-Xuan L. Emotions and Information Diffusioonn Social Media-Sentiments of Microblogs and sharing behavior. *Journal of Management Information Systems* 2013;29(4):217–248.
- [12] Cayon E, Thorp S, Wu E. Immunity and infection: Emerging and developed market sovereign spreads over the Global Financial Crisis. *Emerging Markets Review* 2018;34:62–174.
- [13] Eichler S. The political determinants of Sovereign bond yield spreads. *Journal of International Money and Finance* 2014;46:82–103.
- [14] Calvo G. Servicing the public debt: the role of expectations. *The American Economic Review* 1988;78:647–661.
- [15] Calvo GA, Mendoza EG. Rational Contagion and the globalisation of the securities markets. *Journal of the International Economy* 2000;51:79–113.
- [16] ; 2018. <https://businesstech.co.za/news/finance/233199/the-highest-income-tax-rates-in-the-world-including-south-africa/>.
- [17] peter Vanham, Top 10 things to know about the Mexican Economy;. <https://www.weforum.org/agenda/2015/05/top-10-things-to-know-about-the-mexican-economy/>.
- [18] of Pensilvania WU, What's behind the Volatility of the Mexio's Peso?; 2016. <https://knowledge.wharton.upenn.edu/article/whats-behind-the-volatility-of-mexicos-peso/>.
- [19] Britannica E, Resources and Power;. <https://www.britannica.com/place/South-Africa/Resources-and-power>.
- [20] Bank TW, he Country with theWorlds worst inequality is...;. <https://www.npr.org/sections/goatsandsoda/2018/04/02/598864666/the-country-with-the-worlds-worst-inequality-is>.
- [21] James G, Witten D, Hastie T, Tibshirani R. An Introduction to Statistical Learning. Springer; 2013.